WARWICK
THE UNIVERSITY OF WARWICK

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:
http://wrap.warwick.ac.uk/104207

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

Cancer gene mutation detection in circulating cell-free DNA in blood

by

Hugh Kikuchi BSc MPhil

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Medical Sciences

Division of Biomedical Sciences, Warwick Medical School, University of Warwick

April 2018

Table of Contents

Abbreviations

%	Percent
ΔCt	Delta Cycle threshold
AKT	Alternative name for PKB (Protein Kinase B)
ALK	Anaplastic Lymphoma Kinase
ARMS	Amplification Refractory Mutation System
ATC	Anaplastic Thyroid Carcinoma
BAL	Bronchoalveolar lavage
BEAMing	Beads Emulsion Amplification Magnetics
Bi-PASA	Bidirectional PCR Amplification of Specific Alleles
bp	Base Pair
BRAF	v-Raf murine sarcoma viral oncogene homolog B
BRISQ	Biospecimen reporting for improved study quality
castPCR	Competitive allele-specific Polymerase Chain Reaction, also referred to as CAST
CD340	Cluster of Differentiation 340
CE-IVD	CE marked- In Vitro Diagnostic
CE-SSCA	Capillary Electrophoresis Single-Strand Conformation analysis
cfDNA	Circulating/ Cell Free Deoxyribonucleic Acid
CNG	Copy Number Gain
CNV	Copy Number Variation
COLD-PCR	Co-amplification at Lower Denaturation temperature-PCR
COSMIC	Catalogue of Somatic Mutations in Cancer, online database
CoV	Coefficient of Variance
CR	Conserved Regions
C-Raf	RAF proto-oncogene serine/threonine-protein kinase

CRC	Colorectal Cancer
CRUK	Cancer Research UK
Ct	Cycle Threshold
CT	Computed Tomography
CTC	Circulating Tumour DNA
ctDNA	Circulating Tumour Cell
dH_2O	Distilled and autoclaved water
DLBCL	Diffuse Large B-Cell Lymphoma
DNA	Deoxyribonucleic Acid
dPCR	Digital Polymerase Chain Reaction
dsDNA	Double Stranded Deoxyribonucleic Acid
EBUS	Endobronchial ultrasound
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Rector
EQA	External Quality Assessment
ErbB2	Alternative name for HER2
ErbB3	Alternative name for HER3
EUS	Endoscopic ultrasound
FDA	Food and Drug Administration
FFPE	Formulin Fixed Parafin Embedded
FFPET	Formulin Fixed Parafin Embedded Tumour
FNA	Fine Needle Aspiration
FRET-PCR	Fluorescence Resonance Energy Transfer PCR
g	Gravity

gDNA	Genomic Deoxyribonucleic Acid
GDP	Guanosine diphosphate
GF	GeneFirst
GTP	Guanosine-5'-triphosphate
HE	Hematoxylin and Eosin
HER	Human Epidermal Growth Factor Receptor
HER 3	Human Epidermal Growth Factor Receptor 3
HER2/neu	Human Epidermal Growth Factor Receptor 2/ proto-oncogene neu
HRM	High Resolution Melt
HVR	Hypervariable Region
IHC	Immunohistochemistry
IL	Interleukin
KRAS	Kirsten Rat Sarcoma viral oncogene homolog
LBC	Liquid Based Cytology
LCM	Laser Capture Microdissection
LINE-1	Long interspersed nuclear elements 1
LNA	Linked Nucleic Acid
LoD	Limit of Detection
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization Time Of Flight
MAP	MIDI Activated Pyrophosphorolysis
MAPK	Mitogen-Activated Protein Kinase
MDT	Multidisciplinary Team
ME-PCR	Mutant Enriched Polymerase Chain Reaction
MET	MET Proto-Oncogene, Receptor Tyrosine Kinase, also referred to as c-MET
MGB	Minor Groove Binder

MIDI	Micro-Insertions, Deletions, Indels
miRNA	Micro Ribonucleic Acid
MM	Master Mix
MPE	Malignant Pleural Effusion
MRI	Magnetic Resonance Imaging
MS	Mass Spectrometry
n	Number
NAEDI	National Awareness and Early Diagnosis Initiative
ng	Nanogram
NGS	Next Generation Sequencing
NHS	National Health Service
NICE	National Institute of Health and Care Excellent
NPA	Negative Percent Agreement
NPC	Nasopharyngeal Carcinoma
NPV	Negative Predictive Value
NRAS	Neuroblastoma RAS viral oncogene homolog
NSCLC	Non-Small Cell Lung Cancer
NTC	No Template Control
p	Probability
P.	Page
PAP-A	Pyrophosphorolysis-Activated Polymerisation Allele specific amplification
PBMC	Peripheral Blood Mononuclear Cell
PCR	Polymerase Chain Reaction
PE	Pleural Effusion
PET-CT	Positron Emission Tomography- Computed Tomography

PFS	Progression Free Survival
PGM	Personal Genome Machine (NGS Instrument)
PI3K	Phosphatidylinositol-3-Kinase
PIK3CG	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma isoform
pM	Picomolar
PNA	Peptide Nucleic Acid
PPA	Positive Percent Agreement
PPV	Positive Predictive Value
PTEN	Phosphatase and tensin homolog
Q-PCR	Quantitative Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
RGQ	RotorGene ™ Q (PCR Instrument)
RNA	Ribonucleic Acid
RPO	Residual Polyp of Origin
RT-PCR	Real Time Polymerase Chain Reaction
SAP	Shrimp Alkaline Phosphatase
SD	Stable Disease
SNP	Single Nucleotide Polymorphism
SOP	Standard Operating Procedure
SSCP	Single Strand Conformation Polymorphism
SSLC	Small Cell Lung Cancer
STAT	Signal Transducer and Activator of Transcription
TA	TaqMan Array
Tam-Seq	Tagged-amplicon deep Sequencing
TBNA	Transbronchial Needle Aspiration

TGFα	Transforming Growth Factor Alpha
TKI	Tyrosine Kinase Inhibitor
TLDA	TaqMan Low Density Array (original name for TaqMan Array)
TP53	Tumor Protein p53, also known as P53
UHCW	University Hospital Coventry and Warwickshire
UK	United Kingdom
UKLS	UK Lung Cancer Screening Trial
µl	Microlitre
UV	Ultraviolet
VEGF-A	Vascular Endothelial Growth Factor A
WGS	Whole Genome Sequencing
WT	Wild-type

List of Figures

List of Tables

Acknowledgements

The work presented in this thesis would not be possible without the assistance of a very broad range of people and institutions, who I would like to acknowledge and thank here.

My academic supervisor Prof Ian Cree, whose knowledge and experience are unmatched in the field of molecular pathology and has been a constant source of advice, guidance, support and encouragement throughout the project. My second academic supervisor, Dr Anne Reiman, whose expertise in molecular genetics was invaluable to my supervision, and also delivered a vast majority of my training in the laboratory, gave me the skills to complete this project. My additional supervisor, Prof Dimitris Grammatopoulos, who took over as primary supervisor at the end of the project, was critical to facilitating the completion of the project at Warwick Medical School.

All laboratory work included in this thesis was performed in the Pathology Department at University Hospital Coventry and Warwickshire, so I would like to give my heartfelt thanks to all the staff in the Pathology Department who accommodated me using their laboratory space during their busy diagnostic schedules. In particular I would like to thank the staff of the molecular laboratory who were always so accommodating and supportive: Jenifer Nyoni, Andrew Whitehead, Zandra Woodward, Monika Dutt, Ranveer Chaven, Harriet Pegden, Sarah Clark (extra thanks for proof reading draft 3 of the thesis manuscript), Steve Hamilton and Conor Cairns. The molecular laboratory manager, Tina Wotherspoon, was extremely supportive of our research work and has been instrumental in our assay validations and the constructing of business cases for ongoing work in the department. The clinical scientists: Katie Jones, Megan Goddard, Samantha Horridge and Lisa Hodges, were very supportive and I'd like to thank them for their interest and assistance. In particular, I'd like to thank Lisa

Hodges for her continued support and for proof reading the second draft of my thesis manuscript.

The Arden Tissue Bank was instrumental in obtaining FFPE tissue samples used in this project. I would to thank Sean James the tissue bank manager for his continued support and assistance, as well as the other staff, Andrew White and Adrian Fisk.

The Blood Sciences laboratory of the UHCW Pathology Department was essential for obtaining EDTA blood samples for this project. I would like to thank the Senior BMA, Julie Evans, as well as Mark Huntley and Carl Young for the huge amount of assistance they gave me in terms of understanding the systems in Blood Sciences and how to most efficiently collect the samples.

I would like to thank David Snead (consultant histopathologist) for his support, clinical expertise, insight and interest in our work during this project. Also in Histopathology I'd like to thank Peter Smith (clinical scientist) and Dr William Anderson for their assistance in collecting tissue samples and compiling clinical data. Finally from Pathology, I would like to thank the Deputy Department Head, Dr Judith Timms, who was very supportive throughout the project and was instrumental in bringing our developments into diagnostic use.

Outside of the Pathology Department, I'm very grateful to the clinical staff at UHCW who were instrumental to facilitating the collection of clinical samples for use in this research project. From the lung clinic, I'd like to thank Dr Judith Drought for her active assistance in triaging blood samples to Pathology specifically for research use. From the nursing teams, I would like to thank Lindsay Fitzpatrick, Sam Thomas and Dawn Beaty from the lung nursing team for their support, interest and assistance in identifying appropriate clinical samples for

inclusion in this project. Without the support of the clinical staff this project would not have been anywhere near as complete.

This work involved a number of outside collaborators from Industry. At GeneFirst, I would like to thank CEO Guoliang Fu for initial funding of the project and supply of reagents, as well as senior scientist Michal Bilski for his advice and assistance during the early development phase of this project. At Promega, I would like to thank Fiona Marshall for collaborating with us in the development of the circulating DNA extraction kit, for supplying us with prototype kits, and for continued technical support and advice. At Thermo Fisher Scientific, I would like to thank: Marco Rijnen and Freddie Sharkey from the clinical team for their support of our work; Chris Maddren, Daniel Dancer, Ashley Page and Kicki Bergefall for technical support; Rachel Holding and Charlie Fuller for arranging discounts on consumables; and finally particular thanks go to Andy Gaze as our engineer for all the installation, technical support, troubleshooting and servicing of our sequencing instruments, without whom we'd have (occasionally) been in real trouble!

Beyond the areas of academia and industry, I would like to thank my family and friends who have been utterly critical in supporting me during this degree, and without whom the decision to return to study as a mature student would have been significantly harder. To my fellow students at Warwick Medical School (too numerous to individually name), thank you for the support and interest, and for helping me balance work and life.

Finally I'd thank my parents Greer and Koyu Kikuchi, and my brothers Ian and Tom, for their unwavering support during this very challenging phase of my career, getting to the end wouldn't have been possible without them.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. I, Hugh Kikuchi, declare that it has been composed by myself and has not been submitted in any previous application for any degree and all the research has been undertaken in accordance with University safety policy and Guidelines on Ethical Practice.

I am aware of University regulations governing plagiarism and I declare that this document is all my own work except where I have stated otherwise.

The presented work, including data generation and data analysis, was carried out by the author except in the cases outlined below:

- Tissue sample collection and processing. Tissue samples were collected by UHCW surgical teams as part of their routine delivery of treatment to lung and colorectal cancer patients. Tissue samples were processed and fixed by biomedical scientists in the histopathology laboratory in the UHCW Pathology department. Routine testing of clinical samples for genetic markers was performed by biomedical scientists in the molecular laboratory, the results of which were used to select samples for testing during this project.

- Blood sample collection. Blood samples were collected by phlebotomists in the UHCW Phlebotomy Department. These samples were triaged to the Blood Sciences

laboratory (specimen reception) in the UHCW Pathology department, where they were processed and stored by biomedical scientists and biomedical assistants.

- Prospective testing of cfDNA samples using the validated TaqMan Array assay in UHCW Pathology was performed by BMS staff as part of the routine workflow. The data from these tests are shown in table 3.3.8.

- Statistical analysis and selected figures (Chapter 4). Statistical analysis of data generated in this chapter was analysed by Katherine Lloyd, the statistics expert in our research group. Katherine wrote the methods text describing the analysis done (chapter four methods) and also produced the following figures: 4.3.1.1- 5; 4.3.2; 4.3.3.1- 7; 4.3.7.

List of Publications

Parts of this thesis have been published by the author. The findings from chapter three were published in the journal *Pathogenesis* with the following details:

Authors: Hugh Kikuchi [1,2], Anne Reiman [1,2], Jenifer Nyoni [1], Katherine Lloyd [3], Richard Savage [4], Tina Wotherspoon [1], Lisa Berry [1], David Snead [1,2], Ian A. Cree [1,5,6,*]*Corresponding Author*

Institutions: 1. Department of Pathology – Coventry and Warwickshire Pathology Services (CWPS), University Hospitals Coventry and Warwickshire, Coventry CV2 2DX, UK

2. Warwick Medical School, University of Warwick, University Hospitals Coventry and Warwickshire, Coventry CV2 2DX, UK

3. MOAC DTC, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

4. Systems Biology Centre, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

5. Institute of Ophthalmology, University College London, Bath Street, London EC1V 9EL, UK

6. Centre for Technology Enabled Health Research (CTEHR), Faculty of Health & Life Sciences, Coventry University, Coventry CV1 5FB, UK

Article Title: Development and validation of a TaqMan Array for cancer mutation analysis

Journal details: Pathogenesis 3 (2016), 1-8.

Authors: Lesley Uttley, BSc; Helen Buckley Woods, BSc; Hugh Kikuchi, BSc; Anne Reiman, PhD; Susan Harnan, PhD; Becky L Whiteman, BSc; Sian Taylor Phillips, PhD; Michael Messenger, PhD; Angela Cox, BA, PhD; Dawn Teare, BSc, MSc, PhD; Orla Sheils; Jacqui Shaw; Ian A Cree, MBChB, PhD, FRCPath;*

* *Corresponding author*

Article Title: The evidence base for circulating tumour DNA blood-based biomarkers for the early detection of cancer

Journal details: BMC Cancer 2017 17: 697

Abstract

Background: Lung cancer is the most common cause of cancer death worldwide and is estimated to account for more than 1,380,000 deaths per year. Lung cancer can be separated into two major histological types: Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC), accounting for approximately 15% and 85% of cases respectively. Epidermal Growth Factor Receptor (*EGFR*) is a tyrosine-kinase receptor. In NSCLC *EGFR* overexpression is found in over 80% of cases, and *EGFR* copy number gain (CNG) or amplification is found in nearly 60% of them. Tumours with *EGFR* mutations can be treated using anti-EGFR drugs; however currently genetic analysis has to be performed on tissue which is obtained by biopsy.

Aims: This project aims to investigate alternative methods of obtaining tumour DNA for genetic analysis, to potentially improve or support the current diagnostic process. This project will investigate both new testing methods (molecular assays) and new sources of tumour DNA (cell free DNA from plasma).

Methods: A number of methods were employed during this project. Initially *EGFR* and *KRAS* mutation detection was attempted using a novel Peptide Nucleic Acid Polymerase Chain Reaction (PNA-PCR) assay devised by GeneFirst Ltd (Oxford, UK). The second approach utilised custom designed TaqMan Array 384 well plate assays for the detection of *EGFR*, *KRAS*, *NRAS* and *BRAF* mutations. 40 clinical EDTA blood samples were obtained for the investigation of the use cfDNA for oncogenic mutation detection. Plasma DNA extracted using two automated platforms (Qiagen EZ1 and Promega Maxwell). The extracted DNA was analysed using the Ion Torrent Next Generation Sequencing (NGS) platform.

Results: The GeneFirst novel PNA PCR assays appeared to tolerate low concentration FFPE DNA samples but had a very high false positive rate and the endogenous control assay failed regularly (0- 33.3% failure rate over different assay versions). The TaqMan Array assay was very successful at detecting *EGFR*, *KRAS*, *NRAS* and *BRAF* mutations from FFPE tissue, displaying 97.62% and 94.74% concordance with previously used diagnostic assays (Qiagen Therascreen *EGFR* RGQ PCR and Thermo Fisher *KRAS* castPCR). For the automated isolation of cfDNA, the Promega Maxwell instrument gave consistently superior results to the Qiagen EZ1. CfDNA was successfully used to detect oncogenic mutations using both PCR and NGS assays.

Conclusion: This project has utilised a number of approaches in order to investigate new approaches for the detection of clinically actionable oncogenic mutations, both in FFPE tissue (obtained through surgery or biopsy) and the relatively new cfDNA analyte. Two PCR techniques were compared using DNA from FFPE tissue, and the TaqMan Array assay was shown to be vastly superior. The TaqMan Array was subsequently adopted as the primary diagnostic assay in UHCW Pathology. CfDNA (despite the limited number of samples) showed great potential as an alternative for tissue for detection actionable cancer mutations. The Ion Torrent Next Generation Sequencing system proved to be the most sensitive and powerful technique of the ones utilised here, and will prove an invaluable asset for future development of this work.

Chapter One - Introduction

1.1 LUNG CANCER

Lung cancer is the most common cause of cancer death worldwide and it is estimated to account for more than 1,380,000 deaths a year (Ferlay *et al.*, 2010). In the UK the National Health Service records 39,000 new diagnoses and 35,000 deaths annually (NICE, 2011b). Amongst these cases a large proportion, approximately 90% (Subramanian & Govindan, 2007), are smoking related, however there is still a significant proportion of lung cancer patients who are never smokers. This ratio between smokers and former/non-smokers varies greatly internationally. Gender is also a consideration, as amongst male and female groups of lung cancer patients the proportion of smokers varies. The 90% figure given previous generally applies to North America and Europe, where smoking levels are relatively high. In other areas of world such as Asia, where smoking levels are lower and environmental hazards are more common, they have a much higher proportion of lung cancer patients who have never smoked. For example levels of smoking amongst female lung cancer sufferers in Korea and Hong Kong were previous recorded as 25% (Lee *et al.*, 2000) and 56% (Toh *et al.*, 2004) respectively- far lower than the North American and European levels. Therefore although efforts to reduce tobacco use would have a dramatic effect on lung cancer prevalence, there is still a significant cohort of lung cancers amongst non-smoking patients.

Lung cancer can be separated into two major histological types: Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC). SCLC accounts for approximately 13% of lung cancer worldwide (Woolf *et al.*, 2016). In SCLC the cancer cells contain dense neurosecretory granules. SCLC may have an endocrine or paraneoplastic syndrome association (Rosti *et al.*, 2006), and these cancers are often associated with the larger airways (Leslie, 2011). As the name suggests, the tumour cells in SCLC are much smaller than normal cells, with very little

cytoplasm. Some suggest that the reduced cytoplasm is caused by damage to the control mechanisms within the cell that regulate its size (Leslie, 2011).

The other histological type is Non-Small Cell Lung Cancer (NSCLC). This is the more common type of lung cancer, accounting for approximately 85% of cases worldwide (Ettinger *et al.*, 2010). In England and Wales NSCLC accounts for 72% (NICE, 2012) of lung cancer cases. NSCLC can be further classified into three subtypes: adenocarcinoma, squamous-cell lung carcinoma and large-cell lung carcinoma (Pikor *et al.*, 2013). Adenocarcinoma is the most common form, making up 40% of all lung cancers. Whilst it is often associated with smoking, adenocarcinoma is also the most common form of lung cancer amongst never-smokers (defined as those who have smoked fewer than 100 cigarettes in their life time). Squamous cell carcinoma accounts for approximately 30% of lung cancers and typically occurs in the larger airways. Large cell carcinoma makes up about 9% of lung cancers. They are characterised by large cancer cells, with large and conspicuous nuclei and excess cytoplasm, and may show keratinisation in some cells.

1.1.1 Diagnosis of Lung Cancer in the UK

Early diagnosis is a critical factor in patient prognosis, as the chances of progression-free survival (PFS), stable disease (SD), remission and recovery are maximised if diagnosis is confirmed early and treatment started as soon as possible. As the stage at diagnosis increases, the likelihood of a favourable outcome for the patient reduces dramatically. The challenge for healthcare services is to rapidly test and diagnose those with early symptoms.

In 2008 the National Awareness and Early Diagnosis Initiative (NAEDI) was launched by the Department of Health in collaboration with Cancer Research UK (Richards, 2009) to better inform the public about the symptoms of various cancers including lung, so those showing symptoms can present themselves to their doctors as early as possible. A recent strategy for early detection is the UK Lung Cancer Screening Trial (UKLS) which has offered low-dose CT screening to individuals considered at high risk of developing lung cancer (Field *et al.*, 2016). In total 1994 individuals were scanned by low-dose CT (computed tomography), resulting in 42 (2.1%) new diagnoses of lung cancer. This trial proposes that routine screening of high risk individuals may be an efficient means of detecting early disease and cost effective in terms of treatment. In addition to awareness and early diagnosis, supporting smoking cessation is another key healthcare priority. The NHS Stop Smoking Services were established in 1999 (Dobbie *et al.*, 2015) to support those trying to stop smoking.

Some of the symptoms of lung cancer can often be confused with other less severe conditions, and this may contribute towards late diagnosis in many patients. The main symptom of lung cancer is haemoptysis, but this is not always the first symptom and thus not a reliable symptom for early diagnosis. Haemoptysis can be experienced with or without any of the following other unexplained or persistent symptoms (i.e. lasting longer than three weeks): cough, chest or shoulder pain, dyspnoea, weight loss, chest signs, hoarseness, finger clubbing, cervical/ supraclavicular lymphadenopathy and any features of metastasis (for example in brain, bone, liver or skin) (NICE, 2012).

Upon presenting with any of these symptoms, a suspected lung cancer patient will be referred for a chest x-ray. If this test suggests lung cancer then the patient will be referred

onto a lung cancer multidisciplinary team (MDT), and will be offered a contrast-enhanced Computed Tomography (CT) scan of the chest, liver and adrenals. This is often followed up with a spirometry to assess fitness for definitive treatment (NICE, 2017). Diagnosis requires biopsy, which is also important for staging of the disease by histology and cytology. At this stage the techniques that can be used for confirmation and investigation of the disease depend greatly on the size and location of the lesion. Obtaining tissue may require bronchoscopy, endobronchial ultrasound (EBUS)-guided transbronchial needle aspiration or biopsy, endoscopic ultrasound (EUS)-guided fine needle aspiration or biopsy and non-ultrasound-guided transbronchial needle aspiration (NICE, 2011b). If surgery is considered, then Positron-Emission Tomography (PET) is often used to stage patients more accurately.

If biopsy is possible, the collected tissue is fixed in 10% neutral buffered formalin (4% formaldehyde) and embedded in paraffin wax, producing a Formalin Fixed Paraffin Embedded (FFPE) block of tissue (Cree *et al.*, 2014). Sections of the tumour block are mounted onto microscope slides and are analysed by a histopathologist to classify the tumour and confirm the histology. For patients who are eligible for *EGFR* mutational analysis the neoplastic cell content of the sample must be evaluated, and macrodissection may be performed on the FFPE sample (Cree *et al.*, 2014). This is done to increase the proportion of neoplastic cells in the analysis sample. This material is then used perform mutation testing.

1.1.2 Treatment of Lung Cancer

Patients with a confirmed diagnosis of lung cancer are offered a number of treatments, depending on many factors such as the age and general health of the patient, the stage of the disease and the mutation status of their cancer. Treatment options include: chemotherapy, targeted molecular therapy, immunotherapy, radiotherapy, surgery. As the disease stage increases the treatment options become limited. Radiotherapy and surgery are often not recommended or impossible from stage III onwards as the disease has spread too far to be effectively targeted, or the patient is too ill to consider them.

Currently there is a large range of chemotherapy agents available to clinicians for lung cancer treatment. Guidelines for which compounds to use are complex, and as a result choices of chemotherapy regimen can vary from patient to patient. Standard first line chemotherapy for NSCLC involves a single third generation drug (docetaxel, gemcitabine, paclitaxel or vinorelbine) in combination with a platinum drug, such as carboplatin or cisplatin (NICE, 2011b). The treatment may be changed to monotherapy depending on the patients' tolerance of the combination treatment. For patients with *EGFR* mutation positive NSCLC, then the tyrosine kinase inhibitors (TKIs) gefitinib and erlotinib are recommended for first line treatment, with newer *EGFR* inhibitors reserved for relapse or de novo T790M mutation (NICE, 2011b). For Patients with *EGFR* mutation negative NSCLC, pemetrexed in combination with cisplatin is the recommended first line chemotherapy for adenocarcinomas. For SCLC cisplatin or carboplatin are often given either as a monotherapy or in combination with etoposide (Kalemkerian *et al.*, 2013). The advent of immunotherapy with anti-PDL1 antibody drugs is changing the outlook for lung cancer patients (Lee & Chow, 2014) and may even reach front line consideration. The patient's tolerance of the treatment chosen, incidence

and extent of side effects and the tumour response must be closely monitored, and treatment adjusted accordingly if necessary.

1.2 INTRODUCTION TO COLORECTAL CANCER

Colorectal cancer (CRC) is the second most common form of cancer in women and third most common cancer in men worldwide (El Zoghbi & Cummings, 2016), making CRC the third most common cancer and the fourth most common cause of cancer death worldwide (Favoriti *et al.*, 2016). In the UK every year there are approximately 40,000 new cases (41,112 cases 2013, Cancer Research UK) and approximately 16,000 deaths (16,187 cases 2012, Cancer Research UK). There are a number of lifestyle risk factors associated with developing CRC including: high BMI, lack of physical activity, red/ processed meat consumption, alcohol intake and smoking (Johnson *et al.*, 2013). There is considerable variation in the rates of CRC globally, with approximately 55% of cases occurring in the more economically developed countries (Ferlay *et al.*, 2015). This is mostly likely due to associated lifestyle factors.

Despite the high prevalence of CRC, it also one of the most preventable cancers due to the fact it largely arises from benign lesions (either tubular adenomas or serrated polyps) which later develop into cancers (Grady & Markowitz, 2015). It can take many years for benign adenomas or polyps to become CRC, which provides a considerably larger diagnostic window than many other types of cancer. Therefore one important challenge for CRC prevention is to maximise awareness of symptoms so patients present themselves to their healthcare services during this benign period. Large adenomas and cancers tend to bleed intermittently,

so that fecal occult blood testing is the basis of increasingly successful screening programmes for CRC (HQO, 2009).

The transformation of a benign polyp to colorectal cancer has been studied extensively. A summary of the process is shown in figure 1.2.1, adapted from (Fearon & Vogelstein, 1990). There are a number of genetic alterations that occur during the transformation of a benign polyp into a cancerous lesion. The exact sequence and genes involved varies from case to case, which highlights the need for genetic analysis of tumour DNA during clinical management. Typically the original benign lesion is engulfed by the transformed tumour, however in a small proportion of cases part of the original lesion remains, referred to as the residual polyp of origin (RPO). A logical question in this situation is: are tumours that still contain the original polyp (CRC RPO+) biologically, genetically or clinically different to tumours where there is no residual polyp (CRC RPO-)? This raises the concern that current models (based on CRC RPO- tumours) is not an accurate representation of benign to CRC transformation. In a recent study by Druliner *et al.* (Druliner *et al.*, 2016) 10 RPO+ colorectal cancer samples (CRC RPO +) were obtained and analysed by whole genome sequencing (WGS) and RNA sequencing, and compared to a panel of CRC RPO- tumours. Their analysis revealed that CRC RPO+ tumours are indistinguishable from CRC RPO- tumours. This shows that current models of CRC transformation are representative of both tumour types.

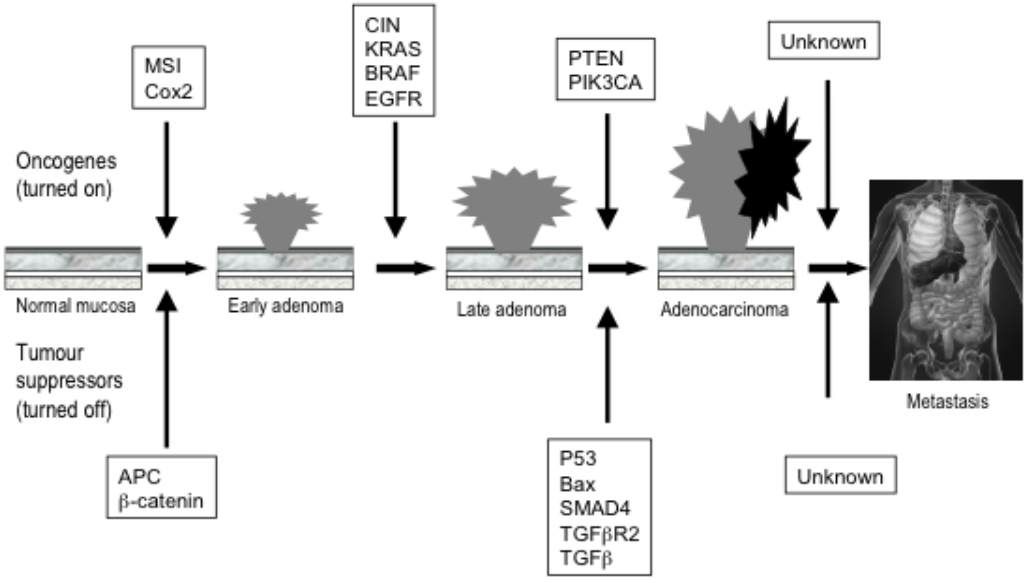

Figure 1.2 Diagram of the major steps involved in benign to CRC transformation, adapted from (Fearon & Vogelstein, 1990). Horizontal arrows show disease progression, vertical arrows and text boxes show the genes typically involved with each stage of disease progression.

1.2.1 Diagnosis of Colorectal Cancer

The four main symptoms of colorectal cancer are abdominal pain, change in bowel habits, loose stools and blood in stools. Upon presentation, investigation will consist of a number of different techniques, including: colonoscopy, sigmoidoscopy with barium enema, CT colonography, or flexible sigmoidoscopy with barium enema (NICE, 2011a). The investigative method used depends on the health of the patient and location of the tumour. If a tumour is identified by either colonoscopy or sigmoidoscopy this will be followed by biopsy for histological classification and genetic analysis. Methods for staging the disease after diagnosis include: contrast-enhanced CT of the chest, abdomen and pelvis; magnetic resonance imaging (MRI); and endorectal ultrasound (NICE, 2011a).

1.2.2 Treatment of Colorectal cancer

For non-metastatic/ local disease, treatment options include chemotherapy (e.g. capecitabine and oxaliplatin), radiotherapy, chemoradiotherapy and surgery (either laparoscopic or open resection). In the event of colonic obstruction, other measures may need to be employed such as applying colonic stents. The exact choice of treatment is typically decided on a case by case basis by the local MDT.

For metastatic disease, treatment frequently involves combination chemotherapy such as oxaliplatin and irinotecan in combination with fluoropyrimidines, or capecitabine and tegafur with uracil (NICE, 2011a). Surgery is considered where possible. Biological agents are also used to treat advanced CRC, including Bevacizumab (VEGF-A inhibitor) or Cetuximab (EGFR inhibitor). Sometimes these agents are combined with chemotherapy agents like oxaliplatin. Another consideration for metastatic disease are secondary tumours, which need to be assessed my MRI or PET-CT and treated as decided by the local MDT.

1.3 GENETIC BASIS OF CRC AND NSCLC

Mutations in the *EGFR-RAS* pathway are common to both CRC and NSCLC, though they differ in the genes most commonly affected, and in the consequences for patients.

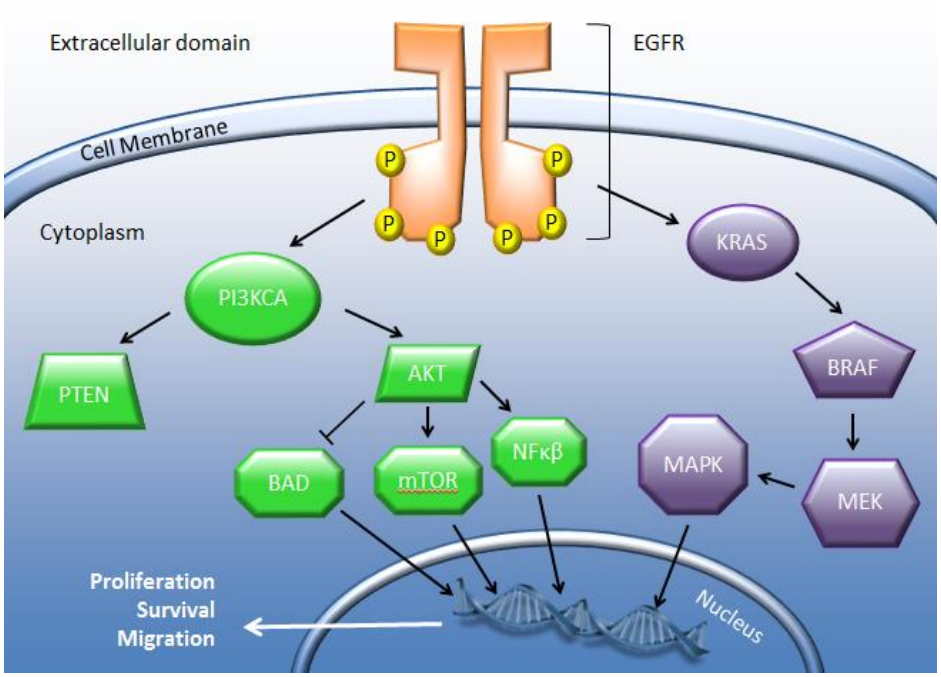

Figure 1.3 Summary of the EGFR and KRAS signalling pathway. Upon activation by its ligand, EGFR initiates a signalling cascade resulting in gene expression in the nucleus, driving the cell towards proliferation, survival and migration.

1.3.1 EGFR

Epidermal Growth Factor Receptor (EGFR) is a tyrosine-kinase receptor from the Human Epidermal Growth Factor Receptor (HER) family of receptors. EGFR is sometimes alternatively referred to as ErbB-1 or HER1. It is a cell surface receptor protein that is constitutively activated by a range of ligands including EGF and TGFα. The HER family of tyrosine-kinase receptors are involved in a number of cellular processes including development, proliferation and survival; all of which are essential for normal cell function, but are also significant in the development of cancer (Jardines *et al.*, 1993).

The other members of the HER family are HER2 (also called CD340, HER2/neu or ErbB2), HER3 (ErbB3) and HER4 (ErbB4). HER2 is similar to EGFR (HER1) as it can heterodimerise with any of the other HER family proteins. In contrast to other HER family proteins, HER2 has no ligand, but heterodimerises with other HER proteins to form active complexes. Amplification of the ERBB2 gene are found in 15-30% of breast cancers (Mitri *et al.*, 2012). HER3 lacks signal transduction capability in a homodimer, but forms active complexes with other HER proteins. HER4 is a single-pass type I transmembrane protein, activated by heparin-binding EGF-like growth factor, betacellulin and neuregulins (NCBI, 2017).

Upon ligand binding EGFR dimerises to form either a homodimer (with another EGFR molecule) or a heterodimer with a related receptor from the HER family (HER2, HER3 or HER4). After dimerisation EGFR autophosporylates tyrosine residues which cause the activation of two main downstream signalling pathways: RAS/MAPK and PI3K/ AKT (Koudelakova *et al.*, 2013). These pathways lead to the cellular processes involved in proliferation and survival (Arteaga, 2003). In normal tissue, these pathways would be activated in the normal regulated manner i.e. cells responding to a biological stimulus. In cancer however, *EGFR* has often been found to contain activation mutations or to have an increased copy number. In NSCLC *EGFR* overexpression is found in over 80% of cases, and *EGFR* copy number gain (CNG) or amplification is found in nearly 60% of them (Koudelakova *et al.*, 2013). Increased *EGFR* activation was previously considered a poor prognostic marker (Hirsch *et al.*, 2003), however a later meta-analysis did not confirm the prognostic significance of *EGFR* overexpression (Nakamura *et al.*, 2006).

The most clinically useful aspect of determining the mutational status of *EGFR* in NSCLC is in assessing how well a patient will respond to tyrosine-kinase inhibitors therapy. The key activating mutations that occur in *EGFR* are exon 21 L858R and exon 19 deletions, which account for approximately 90% of pathogenic mutations (Sholl *et al.*, 2010). Other pathogenic mutations are exon 20 insertions and the (also exon 20) T790M anti-EGFR resistance mutation (Sharma *et al.*, 2007). Being able to detect these specific activating and resistance mutations give clinically actionable information for patient management. Conversely in CRC, *KRAS* mutations have been found to confer resistance to anti-EGFR therapy (Tan & Du, 2012).

1.3.2 *KRAS*

Kirsten Rat Sarcoma viral oncogene homolog (*KRAS*, also known as V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog) is a gene located on human chromosome 12 and encodes the KRAS protein (McGrath *et al.*, 1983). The KRAS protein is a GTPase and thus performs a major role in cell processes such as cell growth, apoptosis and differentiation (Shields *et al.*, 2000). KRAS acts as a molecular on/ off switch, in its GTP-bound form it is active, inducing multiple signalling pathways, and in its GDP-bound form it is inactive. It forms an essential step in many signal transduction pathways downstream of cytokines such as EGF, including C-Raf (Li *et al.*, 2000), PIK3CG (Rubio *et al.*, 1999), MAPK (Vojtek & Der, 1998) and STAT (Vojtek & Der, 1998). The RAS genes were first discovered during the 1960s through research into cancer-causing retroviruses in animals. Experiments using mouse leukaemia viruses resulted in the identification of rat sarcoma genes, and later investigations in the early 1980s

using human cancer cell lines identified human homologs of the rat *KRAS* genes (Pulciani *et al.*, 1982).

The fact that *KRAS* was rapidly identified as an oncogene reflects its significant role in many human cancers. Activating *KRAS* mutations are known to occur in many human malignancies, including approximately 90% of pancreatic cancers, approximately 30% of lung cancers, approximately 60% of thyroid, and approximately 43% colorectal carcinomas (Shackelford *et al.*, 2012). *KRAS* is downstream of *EGFR*, and once EGFR becomes activated GTP bound KRAS will activate its wide range of downstream effectors. Mutations causing constitutive activation of *KRAS* are all point mutations and occur in codons 12 and 13 of exon 2, codon 61 of exon 3, and 117 and 146 of exon 4 (Rajalingam *et al.*, 2007; Cree, 2016b; Cree, 2016a; Sorich *et al.*, 2015). Constitutive activation of KRAS means that all its downstream signalling pathways remain active, giving rise to uncontrolled cell growth and replication, resulting in malignancy.

In NSCLC, *KRAS* mutation most commonly occurs in adenocarcinomas. As yet there has been no clear relationship established between *KRAS* mutation and smoking history, unlike *EGFR* mutations which tend to be more common in non-smokers. As *KRAS* is downstream of *EGFR*, activating mutations in *KRAS* have been shown to result in resistance to anti-EGFR therapy (Shackelford *et al.*, 2012). No RAS mutation dependent therapeutics have yet entered practice. Therefore *KRAS* testing in NSCLC is useful to ensure patients are being treated with the most appropriate regimen and to avoid unnecessary ALK (Anaplastic Lymphoma Kinase) testing. At present the prognostic value of *KRAS* status is not fully understood, however early

detection of *KRAS* mutations would still be extremely valuable for early diagnosis and patient management.

1.3.3 Other significant genes in the *EGFR/ KRAS* pathway

Neuroblastoma RAS viral oncogene homolog (*NRAS*), like *KRAS*, is a member of the RAS gene family. The RAS family are group of GTPases that are involved in a number of pathways regulating cell proliferation, differentiation and apoptosis (Prior & Hancock, 2012). All cells contain three RAS isoforms: *KRAS*, *NRAS* and HRAS (Harvey Rat Sarcoma Virus). *NRAS* was first discovered in 1982 by scientists led by Robin Weiss at the Institute of Cancer Research (Marshall *et al.*, 1982).

The three RAS isoforms share a high degree of sequence homology: 100% homology for a majority of the sequence, with the C-terminal 23-24 amino acids making up the hypervariable region (HVR). This HVR has been shown to be important for the RAS protein membrane interactions and localisation within the cell (Prior & Hancock, 2012) and thus is the most likely explanation for functional and localisation differences between the various isoforms. Figure 1.3.3 shows a diagram of the RAS protein and the degree of homology between the different isoforms across the domains.

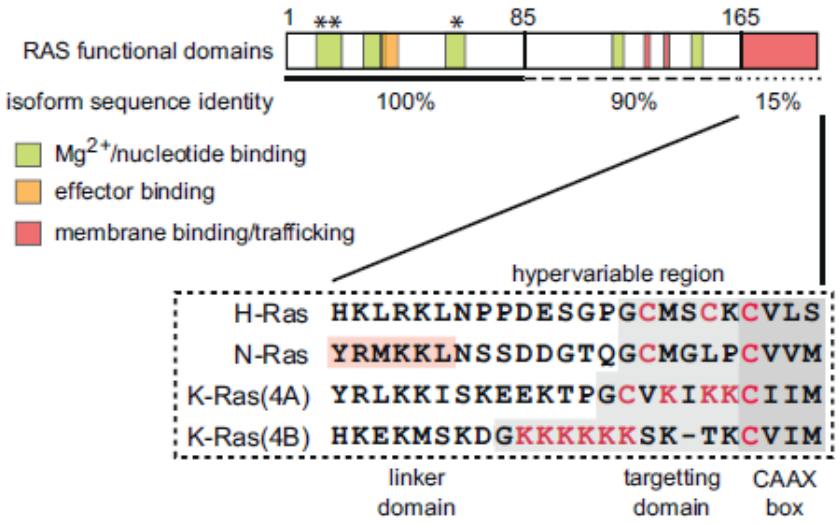

Figure 1.3.3 Diagram of RAS functional domains and differences between isoforms. HVR is post-translationally modified to enable membrane interactions and differential localisation. Asterisks indicate oncogenic mutation sites and codons 12, 13 and 61. Adapted from (Prior & Hancock, 2012).

In the context of cancer, the contribution of *NRAS* is analogous to the mechanism of *KRAS* (described earlier). *NRAS* is most commonly found to be mutated in melanoma, where 20% of cases carry activating mutations in *NRAS* (Milagre *et al.*, 2010). It was the first oncogene to be identified in melanoma in 1984 (Albino *et al.*, 1984). By comparison, *NRAS* mutations are generally found at low abundance in lung and colon cancers. This may be due to *NRAS* being overexpressed in melanocytes compared to the other isoforms, and therefore it is more susceptible to mutation in this tissue type. Clinically the presence of an activating *NRAS* mutation is associated with a poorer prognosis (Devitt *et al.*, 2011). Interestingly there is some evidence that the presence of an *NRAS* mutation may have some correlation with a better treatment response to high dose IL(interleukin)-2 compared to those with wild-type *NRAS* (Fedorenko *et al.*, 2013). At present there are no treatments that specifically target *NRAS*, although are a number are in development. Current clinical trials are largely focused on inhibiting the downstream pathways mitogen-activated protein kinase (MAPK) and

phosphatidylinositol-3-kinase (PI3K), aiming to attenuate the upregulation caused by the *NRAS* mutation (Russo *et al.*, 2014).

The final gene which will be described in this section is *BRAF*, also known as proto-oncogene B-Raf or v-Raf murine sarcoma viral oncogene homolog B. It is a member of the Raf family of signal transduction protein kinases (NCBI, 2016) which are involved in directing cell growth. The BRAF protein consists of 766 amino acids, separated into three domains which are referred to as Conserved Regions (CR) 1-3. Each region is functionally different, with CR1, CR2 and CR3 containing the Ras-GTP binding domain, serine-rich hinge region and protein kinase domains respectively (Daum *et al.*, 1994) (Cutler *et al.*, 1998). *BRAF* mutations are the most commonly identified in melanoma, accounting for approximately 50% of cases (Davies *et al.*, 2002).

Over 90% of cutaneous melanoma samples with *BRAF* mutations contain a substitution of valine to glutamic acid at codon 600, p.V600E (Davies *et al.*, 2002). The high frequency and consistency of mutations in *BRAF* make it an attractive target for therapy. At present there are two main inhibitors used for treating *BRAF* positive melanoma in clinical trials: vemurafenib (Zelboraf) and dabrafenib (Tafinlar) (Wood & Luke, 2016). They were shown to be effective at treating *BRAF* positive melanoma, although there are a number of side effects associated with the toxicity. Another issue with these compounds is they often caused the patients to develop squamous cell carcinomas of the skin (potentially due to paradoxical hyper-activation of the MAPK pathway in wild-type cells), so combination therapy (an anti-BRAF inhibitor combined with a MEK inhibitor) is likely to be further investigated as the optimal treatment regimen (Carlos *et al.*, 2015). Vemurafenib and dabrafenib are currently offered as treatments for *BRAF* positive advanced stage melanoma in the UK.

As with EGFR and KRAS described earlier, NRAS and BRAF are very important targets for characterising patient tumours and identifying the most appropriate treatment option. Whilst *NRAS* and *BRAF* are not the primary focus of this investigation, detection of these genes will be included at certain phases in order to demonstrate that devising accurate molecular techniques can be easily expanded or customised to be appropriate for a wider clinical context.

1.4 NICE GUIDELINES FOR EGFR TESTING IN NSCLC

The National Institute for Health and Care Excellence (NICE) is a UK based organisation that works to create universal comprehensive guidelines and regulations for many aspects of Medicine and Science, with an aim to maximise patient safety and therapeutic effectiveness. Whilst it primarily provides guidance for UK government bodies, the National Health Service and UK Pharmaceutical companies, it also has a large influence internationally. In the UK a majority of healthcare practices design their best practice guidelines to NICE regulations (https://www.nice.org.uk/).

NICE has produced comprehensive guidelines for treatment of lung cancer (described earlier) and also for the use of *EGFR* mutation detection in NSCLC. Clinical trials have shown that lung cancer patients who are positive for *EGFR* activation mutations respond better to tyrosine kinase inhibitor therapy compared to treatment with standard chemotherapy (Hagiwara & Kobayashi, 2013). Therefore it is logical that all patients with NSCLC be tested for *EGFR* mutations (key pathogenic mutations described in section 1.3.1), as this will have a

significant influence on clinical management choices made by their attending physician, ultimately leading to an improved outcome from more appropriate therapy.

The procedure for testing for *EGFR* mutations in NSCLC in UK diagnostic laboratories is not standardised, and there is some variation in the methods used. The methods for testing for *EGFR* mutations can be divided into two key types: targeted mutant detection and mutation screening. With targeted methods, only known mutations are analysed, whereas with mutation screening all known and novel variants are screened for. Many laboratories use both of these types of method for *EGFR* mutant detection. NICE has evaluated many tests currently in diagnostic use in NHS laboratories. Each is summarised in sections 1.4.1 – 1.4.8.

1.4.1 Qiagen™ Therascreen EGFR RGQ PCR Kit

This is a real-time PCR assay that detects 29 *EGFR* mutations (for full list, see section 2.2.6.1). The test first requires DNA to be extracted from FFPE samples using the QIAamp DNA FFPE Tissue Kit, and then a control assay needs to be performed to quantify the total extracted DNA. The Therascreen PCR can then be performed to detect *EGFR* mutants in the sample. To detect the mutant, the Therascreen PCR uses two technologies: Amplification Refractory Mutation System (ARMS) for specific amplification of the mutants; and Scorpions for detection of the amplified regions. For diagrams of the Scorpion-ARMS technology, see section 2.1.

1.4.2 Roche Molecular Systems Cobas® EGFR Mutation Testing Kit

This is another real-time PCR assay for *EGFR* mutations, able to detect 41 different variants. Like Therascreen, the DNA needs to be extracted from the clinical sample using a specific kit, the Cobas® DNA Sample Preparation Kit. The extracted DNA can then be analysed using the Cobas® EGFR assay. The PCR uses complimentary primer pairs and fluorescently labelled probes to detect the mutations. The assay is run using the Cobas® z480 analyser which automates amplification and detection, and the software provides the laboratory with automated test result reporting.

1.4.3 Sanger Sequencing

This method of analysis allows all mutation types to be detected, both known and novel. Sanger sequencing is a very commonly used technique for disease mutant detection, however there is considerable variation in the way the procedures are performed. Typically, DNA will be extracted from the clinical sample and PCR used to amplify the region of interest, in this case, various regions of exons 19-21 of the *EGFR* gene. The amplified fragments will then be cleaned up and sequenced using multiple primers in both directions to ensure adequate sequence coverage and accuracy. A dye terminator cycle sequencing reaction will be used to rebuild the amplified fragments incorporating dideoxynucleotides labelled with fluorescent dyes for sequence determination, and the library of fragments will be analysed using capillary sequencing. The data from this sequencing will be analysed using various software packages to align all the sequence reads and create a consensus for all sequenced regions. Mutations will be identified by comparing the sample consensus with known data

of wild-type and mutant sequences. The advantage of this technique is that it can detect novel mutants; however the disadvantage of this technique is sensitivity in the context of a mixed sample- Sanger sequencing only works well when the tumour DNA makes up at least 25% of the total sample, though this can be improved by careful PCR design: product sizes of <150 bp to take account of DNA fragmentation in FFPE samples (Young *et al.*, 2013).

1.4.4 Pyrosequencing

Pyrosequencing assays are designed to detect specific mutations, and in many labs are used alongside fragment length analysis to identify deletions and insertions. Pyrosequencing works by detecting the release of pyrophosphate molecules upon the incorporation of a specific nucleotide. A typical pyrosequencing protocol is similar to Sanger sequencing: DNA is extracted from the clinical sample and amplified by PCR. The PCR amplicons are then cleaned up and used as the template for the pyrosequencing reactions. This method is more sensitive than Sanger sequencing- the minimum level of tumour DNA in a wild-type sample is 5% (NICE, 2012). An example of one of the pyrosequencing assays available for *EGFR* mutant detection is the Qiagen Therascreen EGFR Pyro Kit (Qiagen, 2011).

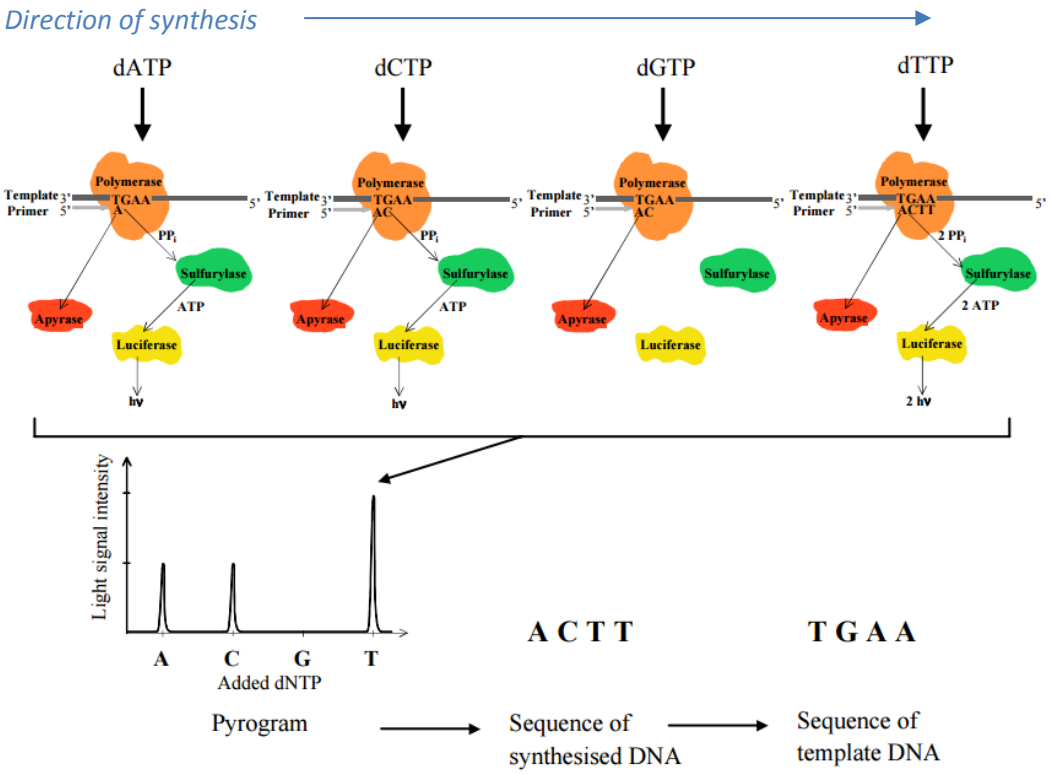

Figure 1.4.4 Diagram representing the chemistry of pyrosequencing technology. With each incorporation of a dNTP molecule, a pyrophosphate molecule is released, which in turn is catalysed by the sulphurylase enzyme to release ATP. The ATP is used as substrate by the luciferase enzyme to generate luciferin and light emission, which is detected by a camera. The strength of the light signal is proportional to amount of ATP present. Adapted from (Ahmadian *et al.*, 2006).

1.4.5 Fragment Length Analysis

Fragment Length Analysis is a technique that distinguishes DNA fragments based on their length, therefore it is best employed for the detection of deletions and insertion events, as substitutions will have no effect on fragment size (Ellison *et al.*, 2013). Extracted DNA samples are extracted, amplified by PCR and fluorescent dyes are incorporated. The

amplicons are then combined with mixed size standards and analysed by electrophoresis. The fluorescence intensity is used to establish the fragment size and detect any insertions or deletions. In some cases, fragment length analysis has been shown to be more effective at detecting exon 19 deletions than direct sequencing (Pan *et al.*, 2005).

1.4.6 Single Strand Conformation Polymorphism Analysis

Single Strand Conformation Polymorphism (SSCP) Analysis is a screening method that distinguishes sequence variations by changes in electrophoretic mobility differences. Extracted DNA is extracted and amplified by PCR. The amplicons are then denatured into single stranded DNA molecules. In single stranded form the DNA will spontaneously undergo three-dimensional folding that will create a unique structure depending on the nucleotide sequence. Single stranded molecules of differing nucleotide sequence will fold differently, and these differences in folding configuration can be assessed by gel electrophoresis (NICE, 2012).

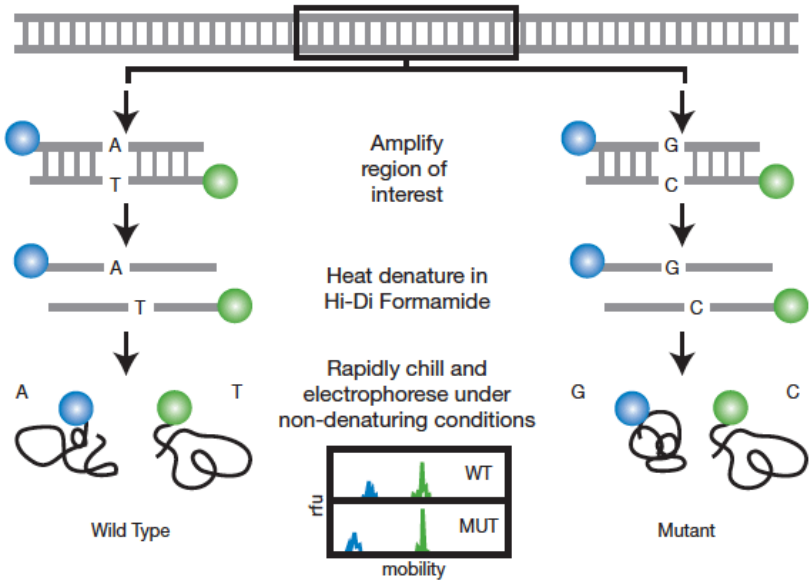

Figure 1.4.6. Diagram demonstrating the principle of Single Strand Confirmation Polymorphism (SSCP) Analysis. Adapted from (ThermoFisher, 2016). Mutant and wild-type single stranded amplicons will three dimensionally fold in a unique conformation which can be distinguished by gel electrophoresis.

1.4.7 High Resolution Melt Analysis

High Resolution Melt (Dubsky *et al.*) Analysis is a fast and cost effective screening method for all types of mutations in double stranded DNA e.g. polymorphisms and epigenetic changes. The method works by analysing the temperature at which different DNA molecules separate as temperature increases and can be combined with COLD-PCR (Co-amplification at Lower Denaturation temperature-PCR) and careful primer design to produce highly specific assays still in clinical diagnostic use (Pichler *et al.*, 2009). Intercalating fluorescent dyes are incorporated into amplified DNA fragments in a PCR step before HRM analysis. The dyes fluoresce strongly when incorporated into double stranded DNA. During the HRM analysis, the increasing temperature causes strands to separate and fluorescence to drop suddenly. By analysing known wild-type samples alongside unknown or mutant samples, the difference in temperature at which fluorescence drops can identify mutant alleles. This

technique can also distinguish heterozygous or homozygous mutants in the analysed amplicons (Pasay *et al.*, 2008).

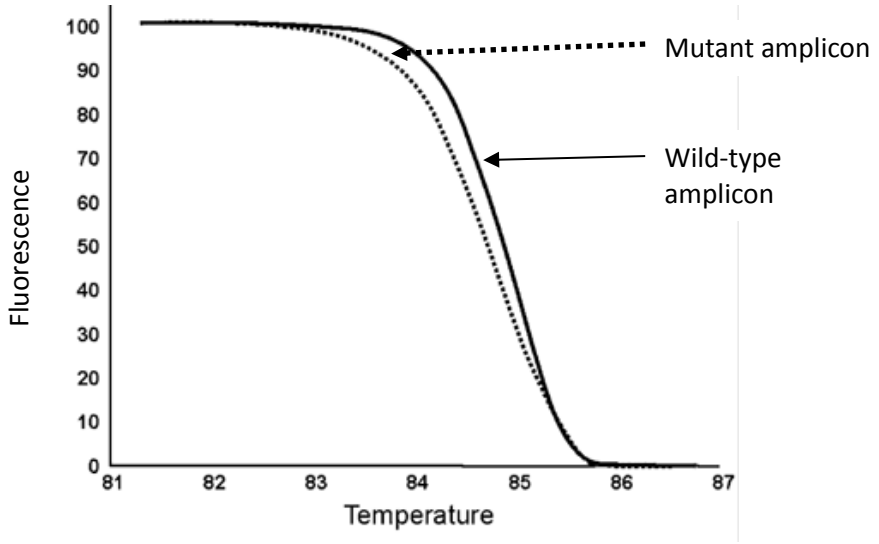

Figure 1.4.7. Diagram showing the melt curves from HRM analysis. The amplicon containing the mutant gives a different melt curve to the amplicon containing the wild-type, which is demonstrated by the change in fluorescence signal. Adapted from (Willmore-Payne *et al.*, 2006).

1.4.8 Next Generation Sequencing

Next Generation Sequencing (NGS) is a relatively new technique, having recently revolutionised genetic analysis options available to molecular researchers. It has by far the broadest and deepest coverage of all the techniques described, but comes with some significant drawbacks. These drawbacks can be summarised as turnaround time and cost. Although NGS platforms vary, generally the library preparation can be very laborious, taking several days to complete before the actual sequence analysis can be run. Related to turnaround time is also the time and skill required for analysis, despite the fact that a lot of

the alignment and variant calling is done automatically, the final analysis can still be very involved, and some expertise is required for the initial setting up of the analysis parameters that the automated systems follow. Cost is also a major issue. The equipment, reagents and consumables for NGS are very expensive, and often only a limited number of patient samples can be analysed in any given run. There are multiplex options available to researchers using NGS systems; however the cost per patient still remains high (Cree *et al.*, 2014) .

1.5 NICE GUIDELINES FOR *KRAS* TESTING IN COLORECTAL CANCER

At present NICE does not offer specific guidance on methods for *KRAS* mutation detection in metastatic CRC. A project to evaluate available methods and produce *KRAS* testing guidance began in 2012, however on 11th September 2014 the project was discontinued (NICE, 2015). The aims of the discontinued project have been combined into a current evaluation of cetuximab and panitumumab (NICE, 2016).

1.6 LIMITATIONS OF EXISTING PROTOCOLS

The limitations of existing protocols and detection techniques can be separated into three categories: firstly the stage at which biological material is available for sampling and testing; secondly the quality and purity of the DNA extracted and analysed once the material is

available; and thirdly the individual assay variability in terms of sensitivity, specificity, limit of detection and scope of mutations they can detect, as well as their associated cost.

In the case of sample collection, due to the non-specific nature of early and mid-stage lung cancer symptoms, investigation and diagnosis do not usually occur until the disease has significantly advanced, thus reducing the patient's favourable prognosis once treatment has begun. Coupled with this the difficulties associated with sample collection make early stage high tumour content samples very difficult to obtain.

Sample quality i.e. the proportion of tumour cells collected for analysis, is highly variable between sample types and can be a major obstacle for mutation detection with some assays. Many molecular assays have high sensitivity and specificity, but may not be able to detect mutant populations in a highly heterogeneous sample. Furthermore certain preservatives, whilst effectively protecting genetic material from degradation, can also cause some alteration to the DNA which can create problems in molecular assays downstream.

1.7 INTRODUCTION TO CELL FREE DNA

Cell Free or Circulating Free DNA (cfDNA) is defined as DNA that has been released from lysed cells. Genetic material is then free to circulate throughout various bodily fluids e.g. blood, pleural effusion, semen and amniotic fluid. The process of releasing DNA occurs naturally during normal bodily processes i.e. apoptosis/necrosis of redundant or dying cells, and is

usually rapidly removed from the blood by the kidneys and liver, as well as being metabolised by plasma nucleases such as DNase1. An alternative mechanism for the release of cfDNA is the shedding of intact cells from tissues, which then go on to lyse and release their DNA content into the surrounding environment. As a result of these natural processes (apoptosis/ necrosis and cell shedding) it is normal for cell free DNA to be detectable in healthy individuals. In a normal healthy subject, cfDNA levels have been recorded as being between 0 and 100ng/ml blood with an average of 30 ng/ml (Anker & Stroun, 2000).

Cell Free DNA is a relatively new marker in molecular diagnostics. The presence of elevated levels of cfDNA in the serum/ plasma of cancer patients was first demonstrated in 1977 (Leon *et al.*, 1977), however it is only recently started to be regarded as a potentially useful diagnostic and prognostic tool. In the context of a cancer patient, the patient's tumour contains a much higher proportion of apoptotic or necrotic cells than normal tissue, and thus mutated tumour cell DNA will be shed into the blood stream, in addition to the background cfDNA released by normal processes. Several studies have quantified higher levels of cfDNA in cancer patients (Shapiro *et al.*, 1983). In addition others have proposed that an increased level of cfDNA in a patient's blood could by itself be used as a diagnostic marker for cancer (Gormally *et al.*, 2007).

The use of cfDNA as a tool for genetic analysis poses a number of technical challenges. Firstly, once the DNA has been released from the protective environment of the cell and nucleus, the DNA is often severely fragmented, with individual molecules rarely being larger than a few hundred base pairs. This makes cfDNA potentially unsuitable for detecting and sequencing large intact gene fragments. This issue may have been recently solved by

Forshew *et al.* (Forshew *et al.*, 2012) who developed a tagged-amplicon deep sequencing (TAm-Seq) method that allowed the full gene analysis of TP53 in plasma samples from ovarian cancer patients. This recent development shows yet more potential for the use of cfDNA for non-invasive molecular analysis of clinical samples.

Secondly, in the case of mutant DNA (e.g. from cancer cells) the total extract of cfDNA from the bodily fluid will contain a heterogeneous mixture of both wild-type tissue DNA and (usually) a much smaller proportion of mutated DNA. This characteristic of cfDNA has been a major technical challenge to investigators seeking to use cfDNA as a diagnostic or prognostic tool, as many technologies require both high concentrations of DNA and a high proportion of mutants in order to detect them, such as conventional Sanger sequencing. In a successful detection assay, the sample must either be enriched for the mutant/ target DNA, and the wild-type DNA removed or suppressed in such a manner so that it does not interfere with the detection of the mutant template.

A key consideration when planning to analyse cfDNA is the sample type that the cfDNA will be isolated from. As described earlier cfDNA can be isolated from a number of bodily fluids, however they are not all equivalent in terms of the quality of DNA that can be obtained from them. In the diagnostic setting, whole blood is readily collected, from which serum or plasma can be derived, so these two sample types are the primary choices as sources of cfDNA for molecular analysis. The study by Lee *et al*. in 2001 (Lee *et al.*, 2001) analysed cell free DNA isolation from plasma and serum, as a source of DNA for detecting post-transfusion chimerism. The findings showed that serum typically contains more cfDNA than plasma, however it is largely genomic DNA from cell lysis during the clotting process (Lee *et al.*, 2001).

This strongly suggests plasma is far more appropriate for isolating circulating tumour DNA. Another study the following year by Thijssen *et al*. (Thijssen *et al.*, 2002) compared cfDNA levels in serum and plasma samples from colorectal cancer patients. Like the paper by Lee *et al.* they found higher cfDNA levels in serum, and serum cfDNA levels were significantly associated with metastasis, whereas plasma DNA levels were predictive for recurrences (Thijssen *et al.*, 2002). Due to the significant amount of genomic DNA in serum, cfDNA from plasma is considered to be more reflective of the amount and composition of cfDNA in vivo. A number of articles and studies utilised cfDNA from both serum and plasma (Xue *et al.*, 2009; Gormally *et al.*, 2007; Aung *et al.*, 2010; Board *et al.*, 2010; Ghorbian & Ardekani, 2012) but did not necessarily compare the relative quality of the cfDNA derived from the two sources. Some studies comparing cfDNA from serum and plasma gave varied results, with some studies finding results from the two samples types were comparable (Morgan *et al.*, 2012) whereas others found plasma to be a better source of cfDNA (Park *et al.*, 2012). In recent articles related to cancer research plasma cfDNA is used most frequently (Zhao *et al.*, 2013; Gonzalez-Cao *et al.*, 2015; Karachaliou *et al.*, 2015).

Beyond cancer, cfDNA has been identified as potentially useful in a number of other contexts. In prenatal testing, cfDNA is being employed for the detection of multiple foetal genetic conditions such as trisomy 21 (Simpson, 2013). In transplantation, donor DNA from donor organs has been detected in recipient blood (Sigdel *et al.*, 2013). CfDNA has been found to have applications in autoimmune disorders such as Systemic Lupus Erythematosis (SLE) (Galeazzi *et al.*, 2003) and Rheumatoid Arthritis (Yan *et al.*, 2006). Increased cfDNA levels have also been described in patients with sepsis (Martins *et al.*, 2000) and has also been reported to occur during strenuous exercise (Breitbach *et al.*, 2012). This range of fields

shows that the study of cell free DNA has a broad range of applications, and development of assays utilising this biological marker may have great clinical utility.

1.8 STUDIES OF *EGFR* MUTATION DETECTION

In order to assess the status of research into *EGFR* testing methods at the time of this project's commencement a literature search was performed. 25 articles were identified and collected by the multiple searches and filtering steps described in the Methods. A summary of all these articles can be found in tables A1.6.1 and A1.6.2 (chapter one appendix). These studies varied greatly in terms of the techniques they analysed, sample types they used, number of samples involved, and the study design. All of the articles were investigating *EGFR* mutation detection methods, using molecular techniques and a majority used clinical samples to assess the properties of the various methods employed.

A wide range of techniques were covered within these articles. Direct sequencing was the most used technique, featuring in sixteen studies, often as a comparator method (Angulo *et al.*, 2012; Betz *et al.*, 2011; Borras *et al.*, 2011; Buttitta *et al.*, 2013; Dahse *et al.*, 2008a; Hu *et al.*, 2012; Kamel-Reid *et al.*, 2012; Liu *et al.*, 2013a; Lopez-Rios *et al.*, 2013; Sriram *et al.*, 2011; Su *et al.*, 2011; Young *et al.*, 2013) but as the primary method in four cases (Billah *et al.*, 2011; Kimura *et al.*, 2006; Smouse *et al.*, 2009; Malapelle *et al.*, 2012). Qiagen Therascreen EGFR assay was also frequently used, being used in five studies (Allegrini *et al.*, 2012; Angulo *et al.*, 2012; Kamel-Reid *et al.*, 2012; Lopez-Rios *et al.*, 2013; Yuan *et al.*, 2012). Fragment size analysis assays (Betz *et al.*, 2011; Kamel-Reid *et al.*, 2012; Su *et al.*, 2011; Young *et al.*, 2013) and High Resolution Melt assays (Borras *et al.*, 2011; Do *et al.*, 2008; Hu *et al.*,

2012; Sriram *et al.*, 2011) were the joint next most popular technique type investigated, each being used in four studies each. Immunohistochemistry (IHC) was used in two studies (Angulo *et al.*, 2012; Liu *et al.*, 2013a). IHC was not the primary type of technique this review focuses on, many other publications using only IHC-related techniques were excluded, however in these two papers other molecular techniques were also being used, and thus the studies were included in collected literature. Pyrosequencing was used in two studies (Borras *et al.*, 2011; Young *et al.*, 2013). NGS was used in only one study (Buttitta *et al.*, 2013).

The sample types used in these studies was also very varied. A majority of studies based their investigations on FFPE tumour samples (either blocks, smears or microscope slide mounted sections), however many other sample types were used, including: Endobronchial ultrasound-guided Fine-Needle Aspirate (EBUS FNA) (Billah *et al.*, 2011), Computed-Tomography guided FNA (CT FNA) (Billah *et al.*, 2011), Ultrasound-guided superficial FNA (Billah *et al.*, 2011), Bronchoalveolar lavage (BAL) (Buttitta *et al.*, 2013), serum/ plasma samples (Hu *et al.*, 2012; Liu *et al.*, 2013a; Sriram *et al.*, 2011; Yuan *et al.*, 2012), cell lines, Malignant Pleural Effusion (MPE) (Kimura *et al.*, 2006; Liu *et al.*, 2013a), Endobronchial ultrasound guided transbronchial needle aspiration (EBUS TBNA) (Schmid-Bindert *et al.*, 2013), and commercially available DNA standards (Young *et al.*, 2013). The numbers of samples used in the studies varied greatly, the lowest was 25 (Yuan *et al.*, 2012) ranging up to 586 (Sriram *et al.*, 2011).

The variety of assays, sample types and numbers of samples encompassed by these articles gave the dataset a considerable amount of depth and scale, with many contrasting studies. Conversely however this very heterogeneous dataset also generated a very variable set of study outcomes. In particular, the studies varied greatly in their aims and study design. Therefore, the degree of equivalence between the studies reviewed was hard to assess.

Some articles described validations of in-house or novel techniques, whilst others used only one technique to analyse different sample types, and others encompassed multiple centres and multiple techniques with a set panel of samples. As a result, many of the studies quantified or qualified their findings using very different measures. Some studies used Positive Predictive Values (PPV) and Negative Predictive Values (NPV), whilst others used sensitivity and specificity, Limit of Detection (LoD) or qualitative measures i.e. mutations detected/ not detected. Some included technical details such as the amount of genomic DNA required to perform the various assays, whilst others did not. This heterogeneity in the dataset has provided a large amount of information, however it also means direct comparisons of some studies is impossible.

Despite the highly varied nature of the articles reviewed, the investigations could be matched together in terms of their study approach. Based on these criteria, the studies could be grouped together into three subsets: studies investigating different tissue sample types for *EGFR* mutation detection; studies investigating different EGFR mutation detection techniques; and studies investigating EGFR mutation detection using Blood or cfDNA.

1.8.1 Studies Investigating Alternative Tissue Sample Types or Sample Processing Methods

Many studies investigated alternative sample types that can be used to obtain tumour DNA for *EGFR* mutation detection. This is a logical approach to improving mutation detection as there are many detection assays available that can achieve high accuracy, and finding a new sample type without having to develop a new assay could prove to be a very cost-efficient and labour-efficient means of improving *EGFR* mutant detection.

The best established sample type for *EGFR* mutation analysis is FFPE tumour material collected by biopsy. As a result, many studies focused on FFPE samples, or used FFPE as a comparator specimen to compare other sample types to. A total of seven studies (Allegrini *et al.*, 2012; Betz *et al.*, 2011; Hu *et al.*, 2012; Liu *et al.*, 2013a; Malapelle *et al.*, 2012; Schmid-Bindert *et al.*, 2013; Smouse *et al.*, 2009; Sriram *et al.*, 2011) included FFPE tumour samples in their investigations. Four studies focused on other sample types resulting in cytological preparations, including Fine-Needle Aspirates (FNA), Bronchoalveolar Lavage (BAL) and Malignant Pleural Effusion (MPE) (Billah *et al.*, 2011; Buttitta *et al.*, 2013; Kimura *et al.*, 2006; Liu *et al.*, 2013a). Studies focusing primarily on FFPE samples will be examined first.

Allegrini *et al.* (Allegrini *et al.*, 2012) used Qiagen Therascreen to test a range of cytological samples including fine needle aspirates (FNA), pleural effusion (PE) or ascitic fluid, and samples from bronchial washing and brushing. This range of samples provided an interesting comparison as the samples are fixed by a number of different methods (ethanol 73.1%, Duboscq-Brasil 16.7%, formalin 9.3%). They analysed the sample types for the entire Qiagen Therascreen panel of *EGFR* mutations. 85.2% of samples were successfully amplified, detecting *EGFR* mutations in 23.9%. Amongst the samples detected positive for *EGFR* mutations, 41% contained less than 200 cancer cells and 18.2% contained less than 50% neoplastic cells. They concluded that cytological samples were a suitable source of genetic material for *EGFR* mutation analysis.

Betz *et al.* (Betz *et al.*, 2011) compared cytological smears against FFPE tumour samples using direct sequencing and PCR-based fragment analysis to analyse the mutation status. Their assays examine *KRAS* codons 12, 13 and 61, and *EGFR* exon 19 deletions and L858R. Interestingly their data showed that cytological smears actually improved mutation

detection over FFPE samples. In dilution experiments *EGFR* mutations were detectable at 10% (proportion of tumour to wild-type), and *KRAS* mutations were detectable at 5%.

Hu *et al.* (Hu *et al.*, 2012) used both multiple sample types and detection methods. They examined three sample types: FFPE tumour samples, frozen tumour samples and serum; and used High Resolution Melting analysis and direct sequencing (as a comparator) for mutation detection. Using these two techniques they analysed EGFR exons 18-21. FFPE tumour samples proved to be the best sample type for mutation detection, as this sample type gave the most positive results (55.67%), followed by frozen tumour samples (51.06%) and serum (46.81%). Compared to direct sequencing, the sensitivity and specificity of the HRM assay was 91.97% and 100% respectively.

The study by Smouse *et al.* (Smouse *et al.*, 2009) aimed to compare the use of FFPE surgical tumour samples against cytology cell blocks for mutation detection. They used only direct sequencing as their analysis method, examining *EGFR* exons 18-21. The study found that 90.9% of the 263 sample analysed were suitable for sequencing. *EGFR* mutations were detected in 28% of FFPE samples and 58% of cytology samples; however cytology samples made up only 5% of the total samples examined in the study. The group concluded that cytological samples were a suitable substitute for tumour samples, giving comparable sensitivity.

Schmid-Bindert *et al.* (Schmid-Bindert *et al.*, 2013) performed a slightly different study to the articles reviewed so far, as they focused on RNA yield for a number of applications, but also included *EGFR* mutational analysis and thus the study was included in this review. They used the TaqMan MGB Real Time PCR assay to detect *EGFR* mutations (exons 19- 21) in three sample types including: endobronchial biopsy, endobronchial ultrasound-guided

transbronchial needle aspirate (EBUS TBNA) and Computed Tomography-guided core biopsy. All three of these sample types were additionally split into two: one aliquot was FFPE preserved, whilst the other was stored in RNAlater until extraction. *EGFR* mutations were detected in 24.8% of samples tested, with EBUS TBNA giving the highest proportion of positive samples (31.3%). All samples came from patients with confirmed NSCLC, however the *EGFR* status had not been previously investigated, so interpretation of these *EGFR* results are unclear.

Malapelle *et al.* (Malapelle *et al.*, 2012) performed another slightly different study. Their investigation examined the use Liquid Based Cytology (LBC) samples like some other studies, but they also analysed effect of a specialist sample processing technique: Laser Capture Microdissection (LCM) of Papanicolaou-stained cells. Direct sequencing was used to analyse the mutation status of *EGFR* (exon 19 deletions) and *KRAS* (codon 12), and the results were compared in samples with and without the use of LCM during sample preparation. The results showed that the use of LCM increased *EGFR* and *KRAS* mutation detection from 21% to 40% by concentrating the neoplastic cell DNA, suggesting insufficient sensitivity of their mutation analysis in standard cytological preparations.

Other studies focusing on sample type for optimising mutations detection used non-FFPE tumour material for their investigations. Billah *et al.* (Billah *et al.*, 2011) used direct sequencing as their analysis method and analysed a range of cytological samples including: endobronchial ultrasound-guided fine-needle aspirates (EBUS FNA), computed-tomography guided fine-needle aspirates (CT FNA), body fluid, ultrasound guided superficial FNA and other cytological specimens. Using direct sequencing they analysed *EGFR* exons 18- 21 and *KRAS* codons 12, 13 and 61. Sample insufficiency was generally low (6.2% overall) ranging

from 3.6% to 33% across the sample types. EBUS samples gave the lowest insufficiency rates. *EGFR* and *KRAS* mutations were detected in 19.4% and 23.6% of samples respectively.

Bronchoalveolar Lavage and Pleural Effusion were the focus of a study by Buttita *et al.* (Buttitta *et al.*, 2013). Samples were previously characterised and grouped as mutant or wild-type. This was the only study to use Next Generation Sequencing (Roche 454) for analysis of *EGFR* (exons 19-21). Direct sequencing was used a comparator. This early NGS method proved to be extremely sensitive, achieving high coverage of the sequencing targets, and very high sensitivity detecting mutant DNA in a 1:10000 dilution with wild-type. In the mutant positive group of samples, NGS detected mutations in 81% of samples, against only 16% detected by Sanger sequencing. In the wild-type group, 42% of samples were identified as positive by NGS. This study showed NGS to be vastly superior to direct sequencing in terms of sensitivity.

Kimura *et al.* (Kimura *et al.*, 2006) also focused on pleural effusion as the primary sample type, as this is a common complication in lung cancer. They extracted the cell free fraction of pleural effusion for analysis, using direct sequencing for mutation detection. *EGFR* mutants were detected in 25.6% of samples. No comparator method was used, as the main focus of the paper was examining the response to gefitinib in *EGFR* mutant positive patients. However, the principle of *EGFR* mutant detection from cell free pleural effusion was demonstrated.

Liu *et al.* 2013 (Liu *et al.*, 2013b) performed a comparison of three *EGFR* mutant detection methodologies: the ADx ARMS kit, Sanger sequencing and mutant-specific Immunohistochemistry; using four different sample types: tumour tissue, malignant pleural effusion (MPE) cell block, MPE supernatant and plasma. The three techniques analysed *EGFR*

exon19 deletions and L858R mutants. Compared to direct sequencing ARMS had a sensitivity and specificity of 81.8% and 100% respectively. Comparing plasma against tissue, ARMS had a sensitivity and specificity of 67.5% and 100%. This investigation showed that tumour tissue was the best source of tumour DNA for mutational analysis, but that that other sample types were also suitable for performing *EGFR* mutations detection.

1.8.2 Studies of *EGFR* Mutation Detection Techniques

EGFR testing in NSCLC has been the focus of research for many groups over the last four years, and as a result there have been a number of recent publications comparing and contrasting the range of detection methods available. This approach can be very complex as assays vary greatly in terms of their sensitivity and specificity, and the range of mutations they can detect. Sanger sequencing was regularly employed in these studies, being the historical 'gold standard' for mutant detection, and many targeted or screening assays were compared against direct sequencing. Interestingly there were also a mixture of commercially available kits used as well as novel and in-house methods.

Qiagen Therascreen again featured heavily in these studies, being a widely available and widely used commercial kit for *EGFR* mutation detection. Angulo *et al.* (Angulo *et al.*, 2012) used Qiagen Therascreen and IHC to analyse *EGFR* in FFPE samples, with direct sequencing as a comparator method. They looked only at exon 19 deletion and L858R. Their results showed that Therascreen was more sensitive and specific than the IHC assay, compared to direct sequencing. Therascreen demonstrated a sensitivity and specificity of 100% (for the two mutations analysed), whereas the IHC assay sensitivity and specificity were 71.4% and 100% respectively. Additionally the Therascreen assay demonstrated a LoD of 5%. The LOD of the IHC assay was not assessed.

Lopez-Rios *et al.* 2013 (Lopez-Rios *et al.*, 2013) compared the COBAS EGFR mutation test against Sanger sequencing and Qiagen Therascreen EGFR kit, all using formalin fixed paraffin embedded tumour samples (FFPET). The COBAS assay is designed for a much wider range of mutations that Therascreen, detecting 41 mutations across *EGFR* exons 18- 21, whereas Therascreen only detects 29 mutations across the same regions. Testing of all their samples using both kits gave COBAS assay a Positive Percent Agreement (PPA) of 98.9% and a Negative Percent Agreement (NPA) of 100% compared to Therascreen. The COBAS PPA and NPA compared to sequencing was 98.8% and 79.3% respectively. They concluded that the COBAS and Therascreen assays were highly reproducible to each other, were far more sensitive than conventional Sanger sequencing. Furthermore COBAS required the least DNA input of all three assays.

Two studies, Kamel-Reid *et al.* (Kamel-Reid *et al.*, 2012) and Young *et al*. (Young *et al.*, 2013) performed larger investigations incorporating multiple testing centres and a larger array of techniques. Kamel-Reid *et al.* prepared mixtures of *EGFR* mutant and wild-type cell lines ranging from 0.1- 50%, which were then distributed along with FFPE tumour slides to each of the five Canadian centres involved in the study. The techniques used by the five centres for analysis of *EGFR* exon 19 deletions and L858R included PCR and fragment analysis, FRET-PCR (Fluorescence Resonance Energy Transfer PCR), RFLP (Restriction Fragment Length Polymorphism), Qiagen Therascreen and direct sequencing. Impressively all centres achieved a 1% LoD, some being able to detect the 0.1% standard using Qiagen Therascreen. For the FFPE samples, the centres achieved a concordance of 87%.

A similar investigation was carried out by Young *et al.* (Young *et al.*, 2013) in the UK. 11 different centres were sent a panel of 15 samples of cell line mixtures of varying mutation

loads (0- 15%) and tasked with assessing the *EGFR* gene for exon 19 deletions and L858R mutations. Seven different methods were used to assess the samples received, including: pyrosequencing, COLD-PCR (Co-amplification at Lower Denaturation temperature-PCR), Sanger Sequencing, CE-SSCA (Capillary Electrophoresis Single-Strand Conformation Analysis), ARMS (Amplification Refractory Mutation System), Qiagen Therascreen, Fragment size analysis. Results showed that all techniques were able to detect exon 19 deletions at ≥5% and L858R at ≥7.5%. COBAS EGFR assay and fragment size analysis (in 3/6 labs) were able to detect exon 19 deletions at 1%. Qiagen Therascreen, ARMS, CE-SSCA and COLD-PCR/ pyrosequencing were able to detect L858R at 1%. Sanger sequencing was found to be least sensitive method used by the 11 centres involved.

High Resolution Melting analysis featured in multiple studies (Borras *et al.*, 2011; Do *et al.*, 2008; Sriram *et al.*, 2011), most likely because it is a fast and cost effective test in comparison to sequencing and commercial PCR methods. Borras *et al.* (Borras *et al.*, 2011) carried out a study of three oncogenes (*EGFR*, *KRAS* and *BRAF*) using HRM and pyrosequencing (*KRAS* only), with direct sequencing as the comparator method. The two techniques were tested using FFPE samples from lung and colorectal cancer patients, targeting five mutations in *EGFR*, six mutations in *KRAS*, and one in *BRAF* (see table A1.6.1, chapter one appendix). HRM detected 100% of mutations in *EGFR*, *KRAS* and *BRAF*, and demonstrated a LoD of 2.5%. Pyrosequencing detected 100% of *KRAS* mutations with higher sensitivity than Sanger sequencing. The study showed HRM to be an effective technique for mutant detection, however as it cannot distinguish between different types of mutant, sequencing will often be required as an additional confirmatory test.

Sriram *et al.* (Sriram *et al.*, 2011) used HRM and Mutant Enriched (ME) PCR (also known as PCR clamping) for detection of *EGFR* exon 19 deletions and L858R mutants, using direct

sequencing as the comparator method. This study involved the largest number of samples of the articles reviewed here: it tested 522 fresh frozen surgically resected tumour samples. It was also one of the few to analyse serum or plasma: 64 of the surgical samples also had matched serum samples, bringing the total number of samples tested to 586. HRM proved to be the slightly more accurate of the two techniques, demonstrating 100% sensitivity and specificity, whilst ME-PCR demonstrated 100% sensitivity and 99% specificity. However, ME-PCR showed better performance with serum than HRM, detecting 50% of mutants in matched sera, whilst HRM only detected 33%. This study clearly illustrated that techniques need to be carefully selected based on the sample types being analysed.

Do *et al.* (Do *et al.*, 2008) also compared HRM against direct sequencing using FFPE samples. The HRM assay was tested using a range of input gDNA concentrations from 1-100 ng. HRM was used to analyse both *EGFR* and *KRAS* mutations. For *EGFR* the HRM technique demonstrated a sensitivity of 100% and a specificity of 88-93% on different exons due to the occurrence of false positives. For *KRAS* HRM demonstrated a sensitivity and specificity of 100%. For both genes the HRM could detect mutants with an input DNA concentration of 1 ng.

Two novel techniques that did not feature in other articles were Bidirectional PCR Amplification of Specific Alleles (Bi-PASA, Dahse *et al.* (Dahse *et al.*, 2008b; Dahse *et al.*, 2008a)) and SNaPShot®, a multiplex PCR system (ThermoFisher, 2012) and a Triplex Sizing Assay were employed by Su *et al.* (Su *et al.*, 2011). In the study by Dahse *et al.*, they designed a novel Bi-PASA assay for exon 19 deletions and a novel allele specific PCR to detect L858R mutations in FFPE adenocarcinoma samples. Direct sequencing was employed as a comparator method. In serial dilution experiments the LoD was 12.5% for Bi-PASA and 25% for the allele specific PCR. This shows a much higher LoD (i.e. poorer analytical sensitivity)

than many other assays assessed in these studies. When testing the adenocarcinoma samples, the assays correctly identified three positive samples.

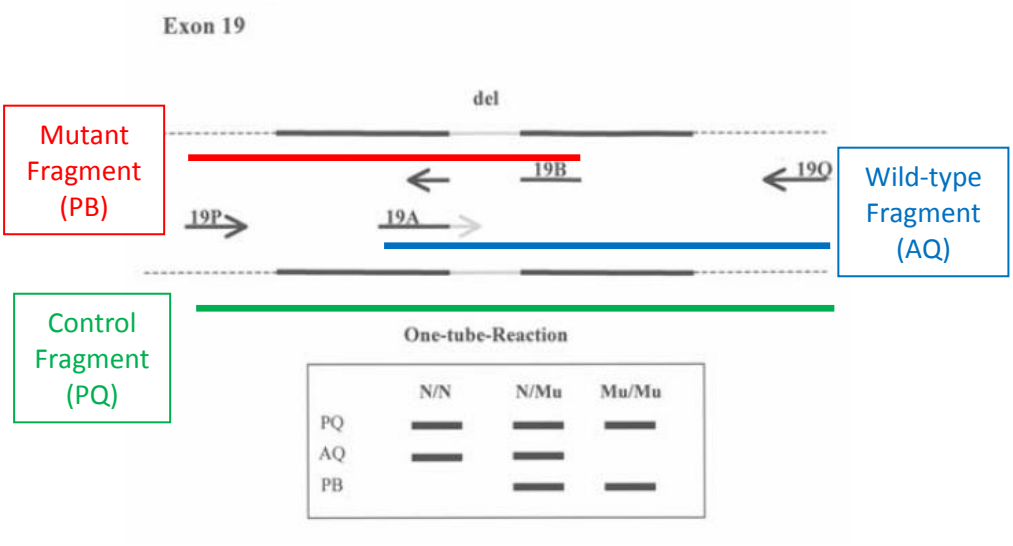

Figure 1.8.2.1 Diagram of the Bi-PASA reaction, in this example covering the *EGFR* exon 19 deletion. Outer primers P and Q flank the mutation site and are always amplified as a reaction control. Inner primers A and B determine the wild-type or mutant status. Primer A is complementary to the wild-type, therefore fragment AQ is only generated by the wild-type. Primer B is complementary to the mutant (deletion), therefore fragment BP is only generated by the mutant. Adapted from (Dahse *et al.*, 2008a).

SNaPShot® and a Triplex Sizing Assay were used by Su *et al.* (Su *et al.*, 2011) for the analysis of multiple cell line and FFPE adenocarcinoma samples. Direct sequencing was used as a comparator. The SNaPShot® technique was able to analyse a broad range of mutations: a total of 38 mutations across eight genes (*EGFR*, *BRAF*, *KRAS*, *NRAS*, *PI3KCA*, *MEK1*, *AKT1* and *PTEN*). The Triples Sizing Assay was used to detect *EGFR* exon 19 deletions and exon 20 insertions, and *HER2* exon 20 insertions. Both assays performed well and detected 100% of mutants in known mutant samples. Both also demonstrated very low LoD: they were able to detect mutations at 1.56%.

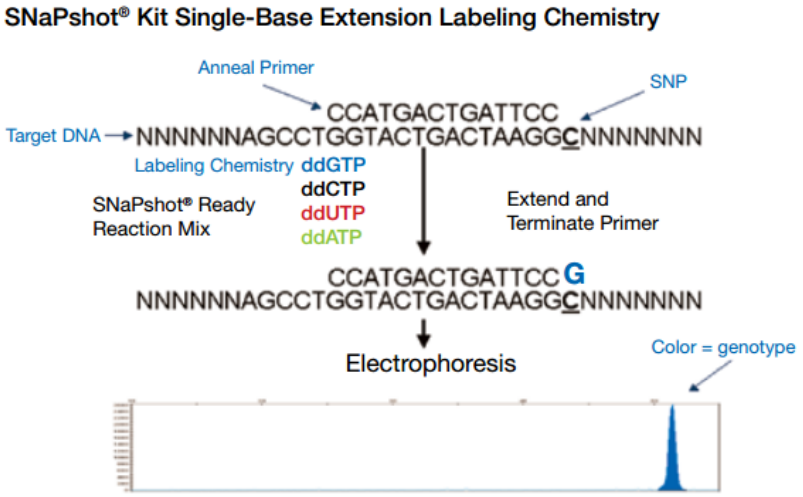

Figure 1.8.2.2 Diagram of the SNaPshot® kit chemistry. Annealing primers bind upstream of the mutation site. Labelled ddNTPs bind to the mutation site which can then be identified by electrophoresis and fluorescence detection. From Life Technologies product brochure (ThermoFisher, 2012).

1.8.3 Studies of *EGFR* Mutant Detection using Blood/ cfDNA

Several PCR based assays have been developed in recent years specifically designed to detect mutations in a small amount of cfDNA with a significant background of wild-type DNA. Whilst there are a range of studies being conducted in this field, relatively few are focusing on *EGFR* mutation detection in NSCLC.

Taniguchi *et al.* (Taniguchi *et al.*, 2011) used BEAMing (Beads Emulsion Amplification Magnetics) to detect *EGFR* mutations in plasma samples from 44 patients. In total 32 (72.7%) out of 44 patient samples were detected as positive for *EGFR* mutations. The properties of BEAMing also allow quantification of mutant alleles within the heterogeneous (in terms of tumour and normal tissue-derived DNA) blood DNA samples. The data showed that mutant alleles constituted between 0.1- 1% of the total DNA isolated. The study suggested that the

sensitivity of BEAMing, and its ability to give a quantitative measure of allele frequency make it a promising tool for tumour mutant detection, and may also prove useful for disease progression monitoring.

Punnose *et al.* (Punnoose *et al.*, 2012) performed a study focusing on Circulating Tumour Cells (CTC) and Circulating Tumour DNA (ctDNA) in trial of Pertuzamab and Erlotinib. A total of 41 patients were included in the study, 25 of which also had plasma samples. CTCs were measured using the Cell Search™ platform. *EGFR* mutations were assessed using the Qiagen DXS EGFR assay. In addition, five other genes were analysed: *KRAS*, *BRAF*, *NRAS*, AKT1 and PI3KCA. CtDNA was found to be superior for detecting *EGFR* mutations in comparison to CTC. The *EGFR* and *KRAS* results from ctDNA were completely concordant with matched tumour samples. CtDNA also detected new mutations post-treatment which were not detected using CTCs, demonstrating the potential use ctDNA for disease monitoring.

The study by Chen *et al.* (Chen *et al.*, 2009) employed MAP (MIDI (micro-insertions, deletions, indels) Activated Pyrophosphorolysis) to detect *EGFR* mutations in the DNA extracted from blood and tissue from lung and breast tumours. MAP was developed by the research group, by modifying an existing technique called Pyrophosphorolysis-Activated Polymerisation Allele specific amplification (PAP-A). PAP-A is an ultra-sensitive method for detecting rare DNA mutations. It works by using allele specific primers that are blocked at their 3' end by a dideoxy molecule (Chen *et al.*, 2009). When a mutant allele is present, the blocked primer binds to the mutant DNA, at which point pyrophosphorolysis removes the 3' block and allows extension to occur. Chen *et al.* modified this technique from PAP-A (just described) to MAP by increasing the number of wild-type sequence mismatches in the allele specific primers, increasing sensitivity. They then applied their modified technique to detecting common *EGFR* mutations in a limited number of clinical samples (4 normal lung tissue, 6 breast cancer

patient plasma and 10 healthy blood samples) and screening a large number of control DNA samples (6400). Using this approach, no *EGFR* mutations were detected in the healthy tissue and blood samples, or in the breast cancer patient plasma samples. However, TP53 mutations that were detected at diagnosis were undetectable after treatment in four of the six women, suggesting clinical remission. MAP detected no *EGFR* mutations in the large cohort of control DNA samples, suggesting that somatic *EGFR* mutation is extremely rare.

Kim *et al.* (Kim *et al.*, 2013) performed a study that aimed to assess *EGFR* mutation detection from plasma cfDNA using Peptide Nucleic Acid (PNA)-mediated PCR clamping in the form of the commercially available PNAClamp ™ EGFR Mutation Detection Kit (PANAGENE Inc., Korea). Plasma samples from 60 patients across 11 centres were tested. All patients showed a partial response to gefitinib therapy. Forty (66.7%) of the 60 patients had previous *EGFR* mutation results (Kim *et al.*, 2013), of which 35 were positive. After testing the plasma samples, 10 cases gave positive results for *EGFR* mutations. *EGFR* mutations were no longer detectable after two months of treatment. The data from this study shows that this particular technique is not sufficiently sensitive to detect *EGFR* mutations in cfDNA compared with matched tissue samples.

1.8.4 Recent Validations of *EGFR* testing Methods

The three previous sections (1.8.1- 3) give an overview of the literature covering various aspects of *EGFR* mutation detection up to late 2014 when this project was in its early stages. Since 2014, there have been a number of publications describing validation studies of new diagnostic methods based on many of the techniques previously described. From 2015 to present there are six publications in particular that present interesting studies, some of

which were diagnostic validations. Of the six studies: three used PCR based techniques (Vliegen *et al.*, 2015; Malapelle *et al.*, 2016; Zhang *et al.*, 2015b) , one used pyrosequencing (Xie *et al.*, 2015), one used NGS (Fujita *et al.*, 2015), and one used MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization Time Of Flight) (Su *et al.*, 2016).

Vliegen *et al.* (Vliegen *et al.*, 2015) validated a combined locked nucleic acid (LNA) PCR/ Sanger sequencing assay for the detection of *EGFR* exon 18 mutations (G719A/S/C) and exon 19 deletions in DNA extracted from FFPE tissues samples. This assay consisted of two steps. First the extracted DNA samples were amplified using the LNA PCR assay. The LNA molecules blocked amplification of the wild-type sequences, therefore enriching the amplicons for the mutant alleles (if present). The amplified products were then sequenced using conventional Sanger sequencing. The results from the LNA/ Sanger assay were compared retrospectively to the original diagnostic results using Fragment Length Analysis and The Qiagen Therascreen EGFR assay. The LNA/ Sanger assay demonstrated a high degree of concordance with the Therascreen results, 100% of samples tested on both techniques gave the same results. The assay also displayed a high level of sensitivity with a LoD of 3.13% tumour cell content. The study concluded that the LNA/ Sanger assay achieved equivalent accuracy, whilst being 25-50% cheaper than the Therascreen assay. The disadvantage however was that the LNA/ Sanger assay required more hands on time, and in its present form does not cover all the most common *EGFR* mutations.

The study by Malapelle *et al.* (Malapelle *et al.*, 2016) aimed to address the problem of the difficulties encountered when attempting to detect *EGFR* mutations from cytological smear samples which are small samples and thus contain a very low number of tumour cells (and therefore a low concentration of tumour DNA). Their approached involved the use of a digital PCR (dPCR) technique, designed to detect *EGFR* exon 19 deletions and the L858R mutation.

The dPCR assay only required 2 ng of input DNA, and can detect mutation rates at 0.1% (Zhang *et al.*, 2015a) so therefore should be well suited to low DNA concentration samples. The results from this study showed the dPCR technique to be sensitive and accurate. The dPCR was tested down to 1% in LoD assays and a majority (19/20) of cytology samples tested gave the same results as previous testing. In one case the dPCR detected an additional L858R mutation which was not detected by the previously used assays (Fragment Length Analysis and TaqMan PCR). The additional mutant was confirmed by cloning and sequencing. This study demonstrated that dPCR is a suitable technique for the detection of *EGFR* mutants, and offers a considerable advantage in terms of low DNA requirement and robustness with low DNA concentration/ quality samples. In a previous study, this group reported that 16.2% of the cytological samples received for analysis had to be rejected due to insufficient cellularity (Malapelle *et al.*, 2013). The improved sensitivity of dPCR may reduce this rejection rate, as well as reduce the service turnaround time and assay failure rate.

Zhang et al (Zhang *et al.*, 2015b) performed a study using the SNaPshot® technique to detect oncogenic mutations in a number of genes including *EGFR*, *KRAS* and *BRAF*, as well as *KIT* and *PIK3CA*. This article is slightly different to others reviewed here, as the clinical context of this article is nasopharyngeal carcinoma (NPC), however this does not impact on the technical aspects of mutation detection assays. Also *EGFR* is considered to be an important mutation in NPC, with up to an 80% prevalence of *EGFR* overexpression in NPC primary biopsies (Pan *et al.*, 2008). *EGFR* mutations however are rare in NPC, at 0-1% prevalence (Lee *et al.*, 2006). 70 fresh frozen tissue samples from NPC patients were extracted and analysed using the SNaPshot® assay. The result of mutation detection from these samples was compared to clinicopathologic data. Of the 70 samples, 12 (17.1%) tested positive for the presence for hotspot mutations in at least one of the five oncogenes analysed (3 in *EGFR*). These results were consistent with the findings of a previous study (Jiang *et al.*, 2014). The SNaPshot® assay

passed all assay controls, however there was no comparator method used to compare the mutation results against. No correlation was found between the mutation status and the clinicopathologic data for each patient.

Xie *et al.* (Xie *et al.*, 2015) performed a large study, using pyrosequencing to analyse FFPE tissue samples from NSCLC and CRC patients. In the total the study analysed 494 and 1099 tissue samples for NSCLC and CRC respectively. Extracts from cell lines were also used to perform LoD experiments. The pyrosequencing technique demonstrated the ability to detect 2% mutant alleles in the wild-type background. Sanger sequencing was used to confirm the results from the LoD experiments with cell lines. In the NSCLC samples, the pyrosequencing detected *EGFR* mutations in 176 of 494 (35.6%). In the 1099 CRC samples, 437 samples contained mutations (39.8%). Sanger sequencing was used to confirm the results generated by the pyrosequencing. These results show a very high prevalence of *EGFR* and *KRAS* mutations amongst NSCLC and CRC patients, and that pyrosequencing appears to be a sensitive and accurate means of detecting mutations in FFPE tissue. However the study reported that in samples tested using both Sanger sequencing and pyrosequencing, pyrosequencing detected more mutations, which is most likely due to the relatively low sensitivity of direct sequencing. As a result the pyrosequencing technique used in routine clinical service at the author's resident hospital.

The study by Fujita *et al.* (Fujita *et al.*, 2015) was a validation study of Next Generation Sequencing assay for the detection of oncogenic mutations from FFPE tissue biopsy samples from NSCLC patients. The NGS assay ran on the Ion Torrent platform, utilising the Ion AmpliSeq™ Cancer Hotspot Panel version 2, which is designed to analyse 50 cancer hotspot genes, including *EGFR* and *KRAS*. 21 tissue samples were analysed using the Ion Torrent assay. A cycleave PCR assay (Yatabe *et al.*, 2006) and Sanger sequencing were used to verify

the results. Comparison of the results showed the mutations detected by the NGS platform matched the results from the other assays in all cases. In addition, in one case a low frequency T790M mutation was identified in 53 of 900 reads (5.9%) for that amplicon, possibly demonstrating the emergence of a resistant subpopulation in that patients tumour. This subpopulation would not be detectable by Sanger sequencing and may not have been detectable by the cycleave PCR assay. This study showed the NGS assay for the accurate, sensitive and practical for use in the clinical setting, especially when there is limited material for molecular analysis.

The final study in this section is by Su *et al.* (Su *et al.*, 2016) and is slightly unusual compared to the other studies as the main mutation detection technique used was matrix-assisted-laser-desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) rather than PCR or sequencing. This article is an interesting contrast to other approaches, but still requires the extraction and amplification of DNA and thus has a molecular biology component similar to other studies. Briefly, the procedure involved extraction of the sample DNA, amplification of the target region, treatment with SAP (shrimp alkaline phosphatase), single nucleotide extension with mutation specific probe, then MALDI-TOF MS analysis. The technique was tested and optimised using mutant and wild-type cell lines, achieving a LoD of 1%. Analytical sensitivity and specificity were tested with a panel of clinical samples. Sanger sequencing was used as the comparator method. The test achieved 100% analytical sensitivity and specificity, so the technique went into routine use. In the course of the period covered by the study, 8147 lung adenocarcinoma samples were tested with this technique, of the failure rate was only 0.1% (n= 5). This study shows an interesting alternative approach to *EGFR* mutation detection. The only technical issue presented was that the technique occasionally detect positive mutants in their wild-type PBMC (peripheral blood mononuclear

cell) DNA which was attributed to background signal, which highlights the need to establish accurate thresholds for clinical diagnostic use.

1.9 SUMMARY

From the review of literature presented in section 1.8 it is very clear that there has been considerable interest in finding the most efficient methods for performing *EGFR* mutation analysis on clinical samples. This interest ranges from investigation of new sample types to use with existing mutation detection methods, to developing novel methods for *EGFR* mutant detection to address the short comings of existing methodologies. It is clear that Sanger sequencing, although for a long time considered the "gold standard" for mutant characterisation in tumour tissue, is gradually becoming obsolete in comparison with NGS technologies, as these have vastly superior coverage and sensitivity. At the time of initial literature searching (2013/14) Sanger sequencing was considered the "gold standard" for mutation detection/ sequence determination. By contrast NGS was less widely adopted, possibly due to the cost and hands on time. In later validation studies of 2015/16 (section 1.8.4) Sanger sequencing was still used in a number of cases, however NGS is very likely to supersede this method soon as speed and cost reduce.

Real time PCR assays offer a much cheaper and quicker alternative to NGS for *EGFR* mutation detection, though limited in mutation coverage. Several techniques and kits using real time PCR technology have been examined here, and have shown to demonstrate high sensitivity, specificity and reproducibility in the studies analysed. Real time PCR also represents a considerable cost saving over more advanced techniques, and can be performed on

machinery that is widely available and affordable. The weaknesses of these approaches are that they lack the sensitivity and range of mutations that other techniques can achieve. The majority of studies use Therascreen or COBAS, and these have also been used as comparators in a number of studies.

FFPE tissue is still the main sample type employed by diagnostic centres for *EGFR* mutation analysis, as DNA yields are consistently the highest from this sample type. Although the use of cfDNA as a source of tumour DNA has previously been limited, work is expanding in this field, and the COBAS (Roche) cfDNA test has recently been approved by the Food and Drug Administration (FDA) (Roche, 2016) and has been used in a number of studies (Weber *et al.*, 2014b; Weber *et al.*, 2014a; Mok *et al.*, 2015) .

This introduction chapter has described the clinical and biological background of lung and colorectal cancer, and the essential need for rapid and accurate means of performing genetic testing on patient tumour samples. The current range of detection techniques have been described in detail, along with some of the newer techniques and approaches that have recently appeared in the literature. The clear underlying problem remains the reliance on tissue which contributes to a significant delay in terms of time between diagnosis and genetic characterisation. This project aims to address these problems, by investigating new methods and new sources of tumour DNA.

1.10 HYPOTHESIS/ RESEARCH QUESTIONS

I suggest that detection of mutations in circulating free DNA in the blood plasma of NSCLC cancer patients is likely to overcome some of the limitations imposed by reliance on tissue in three areas:

(1) Known heterogeneity of mutations within tumours.

(2) Lack of tissue availability from patients.

(3) Development of new mutations during tumour progression.

1.11 PROJECT AIMS

- To compare the new and prototype PCR technology with existing methods for the detection of *EGFR* mutations in blood and formalin-fixed paraffin embedded (FFPE) tumour samples.
- To determine the relative accuracy of new and existing technology for the detection of *KRAS* and *EGFR* mutations in plasma.
- To investigate the stability of cfDNA in EDTA Blood over varying storage times, and the effect of storage time on downstream applications.
- To develop and validate NGS and PCR solutions for blood mutation analysis in lung cancer.

1.12 JUSTIFICATION OF RESEARCH DESIGN

The research project design was initially based on parallel testing of the novel GeneFirst technology against current standard methods for *EGFR* mutation detection in NSCLC samples. The two key variables in the study are firstly; the type of molecular assay employed, and two; the sample type used to obtain the analyte DNA.

The Pathology Department at University Hospital Coventry and Warwickshire NHS Trust (UHCW) is an ideal location for performing this type of research. The laboratories and equipment are all run and maintained to UK Accreditation Service (UKAS) ISO15189 standards, and thus all investigations and evaluations carried out by this project were run to the same standards, and could provide data for analytical and clinical validation to permit rapid introduction of the resulting assay methods to clinical practice. The Department also regularly processes tissue samples for *EGFR* and *KRAS* mutation analysis, and therefore is a good source of reference and comparison material for this project.

Beyond the Pathology Department, UHCW runs a wide variety of clinical services, including regular lung clinics for the treatment and investigation of patients with thoracic conditions. Amongst these patients there were a significant cohort of lung cancer suffers who were undergoing routine blood testing at various points during their treatment. These samples proved to be a good source of plasma for cfDNA extraction and analysis, using standard phlebotomy and processing methods.

A1 CHAPTER ONE APPENDIX

LITERATURE SEARCH

The MEDLINE database was searched using the search tool within the EndNote X7 software on 24th November 2013 to identify original study articles that analysed the currently available methods for detecting *EGFR* mutations in samples from NSCLC patients. Due to the nature of this review, multiple searches were performed in order to capture all relevant articles relating to all the topics within the scope of the article. The details of searches and terms used are given below:

A1.1 Search One- To identify articles related to *EGFR* detection from tissue

(Any field) mutation detection methods lung

[AND] (Title) EGFR mutation lung method*

[OR] (Any field) EGFR mutation testing lung

[OR] (Title) EGFR mutation cytology

A1.2 Search Two- To identify articles related to *EGFR* detection from blood/ cfDNA

(Any field) mutation detection method* lung cell circulating free DNA

[AND] (Title) EGFR mutation lung method* blood

[OR] (Any field) EGFR mutation testing lung cell circulating free DNA

[OR] (Title) EGFR mutation blood

A1.3 Search Three- To identify articles performing direct comparisons of techniques

(Any field) mutation detection methods lung comparison

[AND] (Title) EGFR mutation lung method*

[OR] (Any field) EGFR mutation testing lung comparison

[OR] (Title) EGFR mutation cytology

A1.4 Search Four- To identify articles specifically investigation Cell Free/ Circulating Tumour DNA

(Any field) mutation detection method* lung cell free DNA

[AND] (Title) EGFR mutation lung method*

[OR] (Any field) EGFR mutation lung circulating tumour DNA

[OR] (Title) EGFR mutation circulating tumour DNA

A1.5 Search Five- To identify recent validation studies to update literature search

(Any field) EGFR mutation detection validation

[AND] (Year) 2015

[OR] (Year) 2016

Search results from both searches were filtered by the following criteria:

- English Language
- Non-review article*
- Non-clinical
- Techniques assessed were molecular i.e. designed around the extraction and/ or amplification of DNA.
- Study performed using a minimum of 20 samples.

* *Review articles were not automatically excluded, but were used as sources for identifying original studies. They were not used as primary sources of information.*

Articles that did not meet these criteria were excluded. Search 1 initially yielded 246 articles which were reduced to 90 after filtering. Search 2 initially yielded 244 articles which were reduced to 106 after filtering. Search 3 initially yielded 32 articles, reduced to 9 after filtering. Search 4 produced 18 articles, reduced to 7 after filtering. Search 5 generated 21 Due to the similar terminology in all four searches; there was some duplication of a proportion of the resultant articles. However, doing multiple searches ensured that specific articles were not missed by the individual search algorithms.

A1.6 Literature Results

The abstracts from then articles identified by the five searches were then reviewed for relevance to this study and any that were still not directly relevant to the search criteria were also excluded. To exclude duplicates, the four lists of articles generated by the initial searches, filtering and abstract analysis were combined into a single list and all duplicated entries excluded. After review of the abstracts the numbers of relevant articles from the MEDLINE searches were filtered to 35 full text articles. These articles were reviewed and ten full text articles were identified as irrelevant and excluded. This gave a final citation list of 25 articles.

Table A1.6.1 Articles selected after literature search and filtering. Part 1 of 3 (9 articles), results divided in three for ease of reading. Parts 2 and 3 of table shown in table A1.6.2 and A1.6.3.

Authors	Mutation Testing Method(s)	Comparator Method(s)	Sample type(s)	Mutations Assesed	Number of samples	DNA input required	Findings	Comments
Allegrini *et al.* 2012	Qiagen Therascreen	None	Cytological specimens, including: Fine needle aspirates (FNA), pleural and ascitic fluids, and samples from bronchial brushing and washing.	Exon 19 Deletions, Exon 20 Insertions, G719X, S768I, L858R and L861Q.	108	Not stated	EGFR mutations detected in 23.9% of samples.	Found that EGFR detection is possible from a wide range of cytological samples, although some sample types were more suitable than others.
Angulo *et al.* 2012	Qiagen Therascreen and immunohistochemistry	Direct Sequencing	FFPE tumour samples from NSCLC patients.	Exon 19 Deletions and L858R	136	Not stated	Therascreen detected EGFR mutations to 5% dilution consistently, detected mutations at 1% in half the samples tested; IHC assay sensitivity was 71.4% and specificity of 100%.	Therascreen found to be more sensitive than IHC, compared to direct sequencing.
Betz *et al.* 2011	Direct Sequencing (KRAS and EGFR) and PCR-based Fragment Analysis assay (EGFR).	Compared FFPE cell blocks against direct smears from FNA material as sources of DNA for mutational analysis	FFPE blocks and Cytological smears	KRAS: codons 12, 13 and 61; EGFR: Exon 19 Deletions and L858R	33	Not stated	Direct Smears showed improved DNA quantity, quality and mutation detection rate in comparison to FFPE blocks. EGFR mutants detected at 15%; KRAS mutants detected at 5%.	
Billah *et al.* 2011	Direct sequencing	Compared different sample types using a common mutation detection method	Endobrochial ultrasound-guided (EBUS) fine-needle aspiration (FNA); computed tomography (CT)-guided FNA, body fluid, ultrasound-guided superficial FNA; cytological specimens	EGFR exons 18 to 21; KRAS codons 12, 13 and 61.	99 EBUS FNA, 67 CT FNA, 27 body fluid, 10 ultrasound guided FNA, 6 cytology samples.	Not stated	Sample insufficiency rates varied from 3.6% to 33% between samples types, EBUS specimens gave the lowest insufficiency rates. 88.7% of samples could be sequenced, mutation rates in EGFR and KRAS were 19.4% and 23.6%.	Sample cellularity had a highly significant effect on sample sufficiency.
Borràs *et al.* 2011	High Resolution Melting (HRM) assay and pyrosequencing (KRAS only)	Direct Sequencing	FFPE tumour samples (lung and colorectal)	EGFR: exon 19 deletions, exon 20 insertions, T790M, L858R, P848L; KRAS: G12D, G12A, G12C, G12V, G12S, G13D; and BRAF: V600E.	120	HRM 5- 70ng; pyrosequencing 0.3-1.5ug.	HRM detected 100% of mutants in KRAS, BRAF and EGFR in all samples. HRM was able to detect 2.5% mutant alleles in dilution experiments. Pyrosequencing detected all KRAS mutations from confirmed samples, with greater sensitivity than direct sequencing.	HRM was shown to have high sensitivity and specificity, and could detect mutants at 2.5%. Additionally HRM could detect mutations in multiple genes unlike other kits e.g. Therascreen.
Buttitta *et al.* 2013	Next Generation Sequencing (Roche 454)	Direct Sequencing	Bronchoalveolar lavage (BAL) and pleural fluid. Matched tissue samples had been previously characterised for reference.	EGFR Exon 19 and 21.	33 (BAL) and 15 Pleural effusions	10ng	NGS could detect mutations as low as a 1:10000 diluted in wild-type DNA, and detected mutations in 77% of cytological samples identified as wild-type by Sanger sequencing.	NGS shown to be 1000 times more sensitive than Sanger sequencing for detecting EGFR mutants from BAL and pleural effusion samples.
Dahse *et al.* 2008	Bidirectional PCR amplification of specific alleles; allele specific PCR (ARMS); direct sequencing	None	WT DNA samples from healthy individuals, mutant control cell lines (NCI-H-1650 and H-1975) FFPE adenocarcinoma samples	EGFR Exon 19 Deletions and L858R	35 patient samples	80-100ng	Exon 19 assay detected mutants to a 1:8 dilution (12.5%); L858R assay detected mutants to a 1:4 dilution (25%).	Assays only detects single mutation, Bi-Pasa: Exon 19 Deletions, ARMS: L858R.
Do *et al.* 2008	High Resolution Melting (HRM) analysis	Direct Sequencing	FFPE tumour samples	EGFR Exons 18-21 and KRAS exon 2 codons 12 and 13.	200	1- 100ng (HRM), up to 50ng (Sequencing)	HRM displayed 100% sensitivity but specificity varied from 88-93% on different EGFR exons due to false positives. HRM could also successfully detect mutations with 1ng of template. HRM was 100% sensitive and specific for KRAS mutations.	False positive rate suspected to be due to FFPE DNA alteration.
Hu *et al.* 2012	High Resolution Melting (HRM) analysis	Direct Sequencing	FFPE tumour samples, fresh frozen surgically resected tumour samples and matched serum samples.	EGFR Exons 18- 21.	126 FFPE, 47 frozen tumour samples and 47 serum samples.	Not stated	EGFR mutations were detected at 55.67%, 51.06% and 46.81% in FFPE, frozen surgical samples and serum. HRM sensitivity and specificity was found to be 91.67% and 100%	

Table A1.6.2 Articles selected after literature search and filtering. Part 2 of 3 (8 articles), results divided in three for ease of reading. Parts 1 and 3 of table shown in table A1.6.1 and A1.6.3.

Authors	Mutation Testing Method(s)	Comparator Method(s)	Sample type(s)	Mutations Assesed	Number of samples	DNA input required	Findings	Comments
Kamel-Reid *et al.* 2012	Different methods from five different testing centres were compared, including: PCR and fragment analysis, FRET PCR, RFLP and fragment analysis, Taqman RT-PCR, Qiagen Therascreen, and sequencing.	Various	3 cell lines for optimisation (HCC827, H3255, HT-29), and 30 FFPE tumour samples mounted on slides.	Exon 19 Deletions and L858R.	30	Not stated	All tests achieved a 1% detection rate, with 0.1% being achieved in some locations, and detected cell line DNA with 100% specificity. On tumour samples the sites achieved 87% concordance, however poorer sample quality resulted in some discordant results.	4 of the 5 sites used in house detection methods, only one used a commercial kit (Therascreen).
Kimura *et al.* 2006	Direct Sequencing	None	Pleural Effusion from NSCLC patients	EGFR Exons 18- 21.	43	Not stated	All samples were sufficient for analysis by sequencing, and EGFR mutations were detected in 25.6% of samples.	Sensitivity and Specificity could not be determined as sequencing data from tumour as not available for these patients.
Liu *et al.* 2013	Adx-ARMS kit and IHC.	Direct Sequencing	Tumour tissue, MPE cell block, MPE supernatant and plasma.	EGFR Exon 19 Deletions and L858R.	86 sets of matched samples	Not stated	Compared to ARMS, the sensitivity and specificity og sequencing for EGFR mutant detection was 81.8% and 100%. Compared to tumour tissue, the sensitivity and specificity of plasma using ARMS was 67.5% and 100%.	Whilst tumour is the optimal material for mutation detection, MPE cell block, MPE supernatant and plasma proved to be suitable alternatives.
Lopez-Rios *et al.* 2013	COBAS EGFR Mutation Test	Direct Sequencing and Qiagen Therascreen	FFPE tumour samples	COBAS: 41 mutations in EGFR exons 18, 19, 20 and 21; Therascreen: 29 mutations in EGFR exons 18, 19, 20 and 21; Sequencing: mutation screening in EGFR exons 18, 19, 20 and 21.	124	Test totals (not per well): COBAS 150ng, Therascreen 800ng, Direct Sequencing 600ng.	The COBAS test displayed a PPA of 98.9%, NPA of 100% and OPA 99.2% compared to Therascreen and Sequencing.	Therascreen and COBAS tests are highly concordant. The COBAS test requires the least amount of gDNA of the three tests.
Malapelle *et al.* 2012	Direct sequencing with laser capture microdissection (LCM)	Direct sequencing without laser capture microdissection (LCM)	Cell lines, and liquid based cytology (LBC) samples.	EGFR and KRAS	2 cell lines (PC-9 and H1975), and 42 clinical LBC samples	Not stated	Use of LCM increased EGFR and KRAS mutant detectiont from 21% to 40%.	LCM significantly increases sequencing sensitivity from LBC samples.
Schmid-Bindert *et al.* 2013	Taqman MGB real time PCR assay	Compared different sample types using a common mutation detection method	Forcep biopsy, Endobronchial ultrasound guided transbronchial needle aspiration (EBUS TBNA), and CT-guided core biopsy. Samples were split in two: one was FFPE preserved, the other was stored in RNAlater until extraction.	EGFR Exons 19 and 21.	101	Not stated	EGFR mutations were detecting in 25 of the 101 samples tested.	The results showed that all three samples were suitable for use for molecular analysis.
Smouse *et al.* 2009	Direct Sequencing	Compared different sample types using a common mutation detection method	FFPE tumour samples and cytology cell blocks	EGFR Exons 18-24	263	Not stated	239 (90.9%) samples were sufficient for analysis. Surgical samples yielded lower insufficiency rates than cytology samples (7.4% and 33%). EGFR mutations were detected at 27.8% and 58% of surgical and cytology samples.	Mutations were detectable in sample with 25% or higher tumour cellularity.
Sriram *et al.* 2011	Mutant-Enriched PCR (ME PCR) and High Resolution Melt (HRM) analysis.	Direct Sequencing	Fresh frozen surgically resected tumour samples, and a limited number of matched serum samples. EGFR mutant cell lines (H1650 and H1975) and EGFR WT human gDNA were also used.	EGFR Exon 19 Deletions and L858R.	522 tumour samples, 64 of which had matched serum samples	Not stated	EGFR mutations were detected in 5% of tumour samples. ME-PCR displayed 100% sensitivity and 99% specificity; HRM displayed 100% sensitivity and specificity. In serum, ME-PCR detected 50% of mutants, HRM detected 33%.	Limited detection of only two mutations.

Table A1.6.3 Articles selected after literature search and filtering. Part 3 of 3 (8 articles), results divided in three for ease of reading. Parts 1 and 2 of table shown in table A1.6.1 and A1.6.2.

Authors	Mutation Testing Method(s)	Comparator Method(s)	Sample type(s)	Mutations Assesed	Number of samples	DNA input required	Findings	Comments
Su *et al.* 2011	SNaPshot multiplex PCR assay (Life Technologies); and Triplex Sizing Assay	Direct Sequencing	Cell lines and FFPE	SNaPshot assessed 38 point mutations in eight genes: EGFR, KRAS, BRAF, NRAS, PIK3CA, MEK1, AKT1 and PTEN). Triplex Sizing Assay assessed EGFR Exon 19 Deletions, EGFR Exon 20 insertions and HER2 Exon 20 insertions.	35 lung cancer cell lines, 34 head and neck cell lines, 73 lung adenocarcinomas.	SNaPshot, not stated. Triplex Sizing Assay 1ng.	Both assays identified 100% of known mutant samples. SNaPshot could detct mutants alleles from 12.5% to 1.56% in a WT background. L858R was detected at 1.09%. Triplex Sizing Assay could detect mutants at 6.25% to 1.56% in a WT background.	SNaPshot is limted to known SNPs only, and the Triplex Sizing Assay can only detect deletions or insertions.
Young *et al.* 2013	Seven different methods from 11 different centres were assessed, including: pyrosequencing, COLD-PCR, Sanger Sequencing, CE-SSCA, ARMS, Qiagen Therascreen, Fragment size analysis.	N/A	H1975 cell line (L858R control), and genomic DNA reference standard for Exon 19 Deletions (Horizon Diagnostics, UK)	EGFR Exon 19 Deletion and L858R.	15 samples for each of the 11 participating centres.	Not stated	All centres could detect mutant samples at 7.5%, and 5% was achieved by all but one centre. The most sensitve techniques were Therascreen, ARMS, CE-SSCA and COLD-PCR, detecting the mutants at 1%.	Sanger Sequencing was the least sensitive method. No false positives were reported.
Yuan *et al.* 2012	Dxs EGFR Mutation Kit	Compared their Modified Phenol Chloroform (MPC) extraction method with Qiagen QIAmp MinElute Virus Spin Kit	EDTA plasma	EGFR Exon 19 Deletions	25	Not stated	Results were 100% concordant with previous sequencing results. However only 1/25 samples was a confirmed EGFR mutant, others were all confirmed wild-type.	MPC method yielded CFDNA of higher concentration and lower fragmentation, and gave a stronger CT in mutant detection assay.
Dahse *et al.* 2008	Bidirectional PCR amplification of specific alleles (Bi-PASA)	Direct sequencing	Normal controls and lung adenocarinoma cell lines	EGFR Exon 19 Deletion and L858R.	Not stated	80- 100 ng	The assay successfully detected EGFR mutations in cell lines	
Taniguchi *et al.* 2011	BEAMing (beads, emulsion, amplification and magnetics)	EGFR PNA-LNA PCR assay	Cell free DNA	EGFR Exons 19 and 21.	44 (all EGFR mutant positive)	Not stated	Matching mutants were found using BEAMing in 32/ 44 (72%) of samples	In addition to mutant detection, BEAMing is also able to assess the fraction of mutated alleles in a sample
Punnoose *et al.* 2012	Qiagen DxS (Therascreen) and custom TaqMan assay (exon 20 only)	None	Cell free DNA, DNA from CTC, tissue DNA	EGFR Exons 19 and 21.	41	20 ng	EGFR mutations were found in 4 cfDNA samples, which matched archival tissue results	cfDNA was found to be more successful at detecting EGFR mutations than CTCs.
Chen *et al.* 2009	MIDI-Activated Pyrophosphorolysis (MAP)	Direct sequencing	lung tissue, plasma and blood, and synthetic controls	Various MIDI targets in EGFR and TP53	20 clinical, 6400 controls	Various	MAP demonstrated high sensivity in controls. No EGFR mutations were detected in clinical samples, but TP53 mutations were not detected post-treatment.	Whilst MAP showed good sensitivity, certain analyses had very high DNA requirements.
Kim *et al.* 2013	PNAClamp ™ EGFR Mutation Detection Kit (PANAGENE Inc., Korea)	None	Cell free DNA	EGFR Exon 19 Deletion and L858R.	60	30 ng	10/35 (29%) samples matched previously detected EGFR results from tissue.	Assay proved to not be sufficiently sensitive for detecting EGFR mutations in plasma compared to tissue.

Chapter Two – Detection of *EGFR* and *KRAS* mutants using the novel GeneFirst™ PNA PCR assay

2.1 INTRODUCTION

Prior to 1983, the study of DNA was a relatively crude field in its infancy. In order to isolate enough DNA for analysis, DNA was extracted in relatively high quantities from cells, as there was no way of amplifying the DNA from a small sample. As a result, most DNA research was performed using cell lines or tissue samples, as smaller samples would not yield sufficient DNA for analysis. In 1983 Kary Mullis developed Polymerase Chain Reaction (Mullis *et al.*, 1986), also known as PCR, a technique which would go on to revolutionise molecular biology. Kary Mullis went on to receive the Nobel Prize for Chemistry in 1993 for his work on PCR. This technique involves manipulation of the properties of DNA and the enzyme DNA polymerase, and by cyclically changing the temperature of a mixture of template DNA, DNA polymerase enzyme and primers (short DNA oligonucleotides complementary to the genetic region of interest) the original DNA sample can be amplified by many orders of magnitude. This incredible tool gave scientists the ability to analyse very small samples of DNA, and since its invention PCR has been applied in a broad range of applications including: forensics, genetics, microbiology, and diagnostic medicine.

As with many molecular biology techniques, PCR has developed through many generations since its initial invention, and now there are many different types of PCR designed for different applications. Probably the most significant evolution of the technique is what is commonly referred to as real-time or quantitative PCR (RT or Q-PCR). The first version of PCR, whilst revolutionary, was a qualitative technique, impossible to accurately quantify, and vulnerable to reduced sensitivity and specificity under certain circumstances, including the presence of inhibitors.

Q-PCR was developed in the early 1990s (Holland *et al.*, 1991) (Higuchi *et al.*, 1993), and works on the same basic principle as standard PCR, but with the addition of a fluorescent intercalating dye SYBR green™ which could be detected in real time (Woo *et al.*, 1998). As the technique developed fluorescently tagged molecular probes were introduced as a more specific alternative to SYBR green. These probes are similar to primers (short DNA oligonucleotides complementary to a specific target DNA sequence) but they additionally have a fluorescent "reporter" dye molecule bound to the oligonucleotide. Typically the probes also possess a "quencher", an inhibitor molecule that chemically prevents the fluorescent dye from activating. As the probe is incorporated into the synthesised DNA molecule (during PCR) the quencher is removed resulting in activation of the reporter dye molecule, giving a detectable fluorescent signal (typically under stimulation with UV light or a laser). This evolution of the PCR method gives far greater specificity and accuracy, because the PCR amplification is not measured simply by the amount of DNA amplified but by the fluorescent signal generated by amplification of the specific target region. As a result, the measured amplification is less prone to contamination or false-positives/ negatives, and thus the assay has improved sensitivity and specificity compared to standard PCR. An additional advantage of this technique is it can be quantitative with inclusion of a panel of standards run alongside the samples.

The next evolution of PCR has involved the addition of specific inhibitors to further increase PCR specificity. This has particular significance in the field cancer gene mutation detection from clinical samples. With a clinical sample, for example from a biopsy of a suspected cancerous legion, there will be a high degree of heterogeneity in the tissue recovered. With any given tumour sample, the percentage of neoplastic cells within the tumour tissue may be anywhere between 1 and 100%, meaning the DNA extracted from such a sample will be

similarly heterogeneous. With standard PCR or Q-PCR, it is very likely that the wild-type DNA will outcompete the mutated DNA during amplification, giving a false negative result.

To counter this, new molecular tools have been developed to suppress the amplification of wild-type DNA sequences. Examples include: Scorpions (Thelwell *et al.,* 2000), Peptide Nucleic Acids (Nielsen & Egholm, 1999) and Locked Nucleic Acids (Kumar et al., 1998). These molecules function in a number of different ways, but essentially they all work to suppress the amplification of the wild-type DNA sequences, resulting in the only amplification being from the minority mutant DNA population in the mixed sample. A number of commercial assays utilising this technology are available (see Introduction) two of which (Qiagen Therascreen and Life Technologies castPCR) are in use at UHCW Pathology Department for the detection of cancer mutations as part of their diagnostic service. The Scorpion-ARMS technology incorporated into the Qiagen assay is shown in figures 2.1.1 and 2.1.2 below.

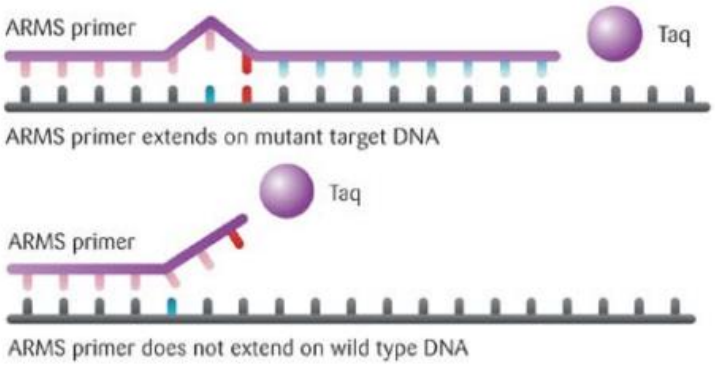

Figure 2.1.1 ARMS primers are designed to match the target mutant DNA sequence. They will not hybridise to the wild-type DNA sequence, and thus Taq polymerase will not extend the DNA strand. From (Berninger, 2010).

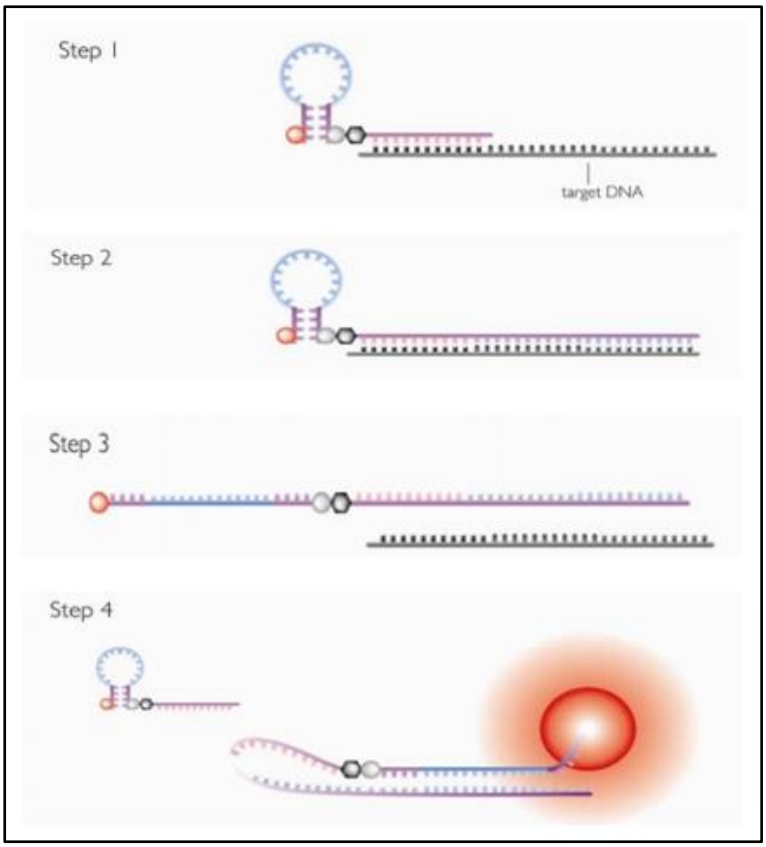

Figure 2.1.2 Diagram showing the mechanism of the Scorpion Primer Probe. Step 1) The Scorpion primer binds upstream of the target DNA region. Step 2) The Scorpion primer is extended by DNA polymerase. The copied region is complementary to the distal end of the Scorpion prime. The blocking group (black) stops DNA polymerase from copying the probe region. Step 3). The newly extended Scorpion primer is denatured by increasing temperature. Step 4) As the PCR reaction cools the extended Scorpion primer rearranges and hybridises to itself. The fluorophore is now no longer in close proximity to the quencher and subsequently begins to fluoresce. Any un-extended Scorpions re-anneal and are quenched. Adapted from (Berninger, 2010).

Peptide Nucleic Acids (PNA) were developed by Nielson and Egholm in 1999 (Nielsen & Egholm, 1999). The name PNA is a misnomer, as technically these molecules are not nucleic acids, but simulate their binding properties. PNA have a very simple structure, with their main obvious difference to nucleic acid is that they do not possess a pentose ring or phosphate groups. Figure 2.1.3 shows the structure of a PNA molecule and a PNA/ DNA complex. The backbone is composed of repeating N-(2-aminoethyl)-glycine units linked by

amide bonds (Nielsen & Egholm, 1999). The bases (A, T, G, C) are attached to the glycine backbone by methyl carbonyl linkages.

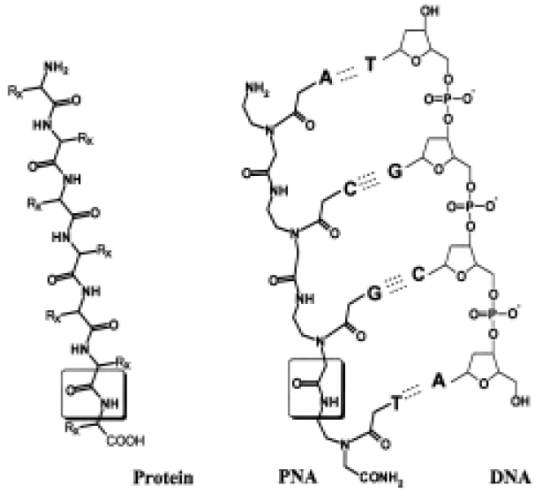

Figure 2.1.3: Right: PNA molecule. Left: PNA/ DNA duplex. From (Nielsen & Egholm, 1999)

The glycine backbone of the PNA molecule means that the PNA backbone has a neutral charge, contrasting with DNA, which is negatively charged. As a result of this there is less repulsion between the binding molecules and therefore PNA binds DNA with a higher affinity than DNA/ DNA complexes (Egholm *et al.*, 1993). Similar data has also been found for PNA/ RNA complexes (Jensen *et al.*, 1997). This higher affinity binding is particular advantageous when using PNA as an inhibitor of DNA replication or transcription. PNA also bind DNA with a greater specificity than DNA/DNA, which has an obvious advantage in terms of assay design for specific targets. Mismatched bases in a PNA/DNA duplex are more destabilising compared to an equivalent DNA/DNA duplex (Nielsen & Egholm, 1999), so this reduces the chance of false positive results from a partially bound PNA oligomers.

GeneFirst Ltd. (Oxford, UK) developed novel PCR based assays for the detection of *EGFR* and *KRAS* mutations from Cell Free DNA. The assays were designed to detect 11 *EGFR* mutations

and 7 *KRAS* mutations with the DNA making up less than 1% of the total DNA sample with wild-type DNA as the majority template. The PCR methods used Peptide Nucleic Acids (PNA) to block the non-specific amplification of the wild-type DNA, so only the mutant DNA is amplified and detected by the assay. Mutant allele amplification is enhanced by the use of Locked Nucleic Acid (LNA) oligonucleotides that bind to the mutant alleles and chemically stabilise oligonucleotide binding.

This technology represented a potential step forward in *EGFR* mutant detection, as it was predicted to be sufficiently sensitive to detect mutations in cfDNA. Collecting a blood sample is a relatively non-invasive procedure, and can be done at the very earliest stage once lung cancer investigation has begun. Thus if the test is successful, tailored treatment can begin immediately without the need for biopsies which may not be obtainable from the patient due to the site of the tumour or co-morbidity.

As a first stage in this investigation, the prototype versions were tested using tissue, as this is the current standard method for detecting cancer mutations. If their effectiveness using tissue could be proven, then this is a fair validation of their performance, and it would then be logical to then move on to testing cfDNA. Another consideration is sample availability. Patient plasma samples taken for research were scarce and of limited volume, hence decisions about how to best use them had to be carefully considered. Beginning testing of these prototype assays with cfDNA would risk wasting precious cfDNA samples until the assay conditions are not yet fully optimised.

2.1.1 AIMS

- To compare the GeneFirst technology with the existing methods (Qiagen Therascreen EGFR RGQ PCR assay and Thermo Fisher *KRAS* castPCR assay) for the detection of *EGFR* mutations in blood, cytology and formalin-fixed paraffin embedded (FFPE) tumour samples. These existing methods (specified above) will be considered the "gold standard" for mutation detection in this phase of the project.

- To determine the accuracy of the GeneFirst technology for the detection of *KRAS* and *EGFR* mutations in plasma versus tissue.

- To determine the compatibility of the GeneFirst assays for use with cell free DNA.

2.2 METHODS

2.2.1 Equipment

Bench top centrifuge compatible with 1.5ml micro-centrifuge tubes (up to 15500 x g)

Bench top centrifuge compatible with 96 well PCR plates (up to 1600g)

Bench top centrifuge compatible with 0.2 ml PCR strip tubes

Bench top centrifuge compatible with Ion Torrent sequencing chips (314v2, 316v2 and 318v2)

Heat block capable of heating 1.5ml tubes to 70°C

Promega Maxwell Automated Extraction Instrument

Laminar Flow safety cabinet with UV decontamination feature

Qubit 2.0 Nucleic Acid Analyser

Life Technologies ViiA 7 Dx Real Time PCR instrument with Fast 96 well block

Life Technologies Ion Chef instrument (see figure 2.2.8.3.2)

Life Technologies Ion Torrent PGM instrument (see figure 2.2.8.3.2)

Gel Electrophoresis Apparatus and UV imaging equipment

Millipore water purification system

2.2.2 Kits

GeneFirst EGFR PCR kit (prototype, versions 1-5)

GeneFirst *KRAS* PCR kit (prototype, versions 1-5)

Qubit dsDNA High Sensitivity reagent kit

Promega Maxwell FFPE DNA Kit

Promega Maxwell Circulating DNA Kit (prototype)

Life Technologies Ion AmpliSeq™ Colon and Lung Panel

Life Technologies Ion AmpliSeq™ Library Kit 2.0

Life Technologies Ion Express Barcode 1-16 Kit

Life Technologies Ion Express Barcode 17- 32 Kit

Life Technologies Ion Torrent PGM Sequencing 200 Kit

Life Technologies Ion PGM IC 200 Kit

Life Technologies Oncomine™ Solid Tumour DNA Kit

2.2.3 Pathology Department Extracted DNA Samples

Previously extracted DNA samples were collected from UHCW Pathology Department. These samples were processed as part of the Department's routine diagnostic services (*EGFR* and *KRAS* testing for lung and colorectal cancer, SOP MM RA 19 and SOP MM RA 20).

2.2.4 Lung Cancer (NSCLC) Extracted DNA samples

FFPE tissue sample were processed for DNA extraction according to the Pathology Department SOP (SOP MM WI 12). Briefly, thin sections from FFPE tumour samples were mounted on microscope slides, stained with Hematoxylin and Eosin (HE) and examined by a histopathologist. The histopathologist confirmed the type of cancer and staged the disease. Samples from patients with confirmed NSCLC and requested for *EGFR* mutation analysis were transferred to the Pathology department molecular team. Here the FFPE sections were scraped from the slides and extracted using the Qiagen EZ1 automated extraction platform

according the manufacturer's instructions (Qiagen EZ1 DSP Virus 2.0 kit). DNA samples were eluted in 60 µl elution volumes and stored at -20°C until ready for molecular analysis.

2.2.5 CRC Extracted DNA samples

FFPE tissue sample were processed for DNA extraction according to the Pathology Department SOP (SOP MM RA 20). Briefly, thin sections from FFPE tumour samples were mounted on microscope slides, stained with Hematoxylin and Eosin (HE) and examined by a histopathologist. The histopathologist confirmed the type of cancer, staged the disease and marked areas of the slide with high tumour content. Samples from patients with confirmed NSCLC and requested for *KRAS* mutation analysis were transferred to the Pathology department molecular team. Here the marked slide from the histopathologist was used to manually micro-dissect high tumour content sections from the original FFPE tumour block using a disposable 1 mm skin punch. These punched FFPE sections were incubated in lysis buffer at 70 °C for up to 4 hours to dissolve the paraffin and lyse the cellular material. The lysate was then extracted using the Promega Maxwell automated extraction platform using the Maxwell FFPE Extraction kit according to manufacturer's instructions. DNA samples were eluted in 60 µl elution volumes and stored at -20 °C until ready for molecular analysis.

2.2.6 UHCW Pathology Department Mutation Results

2.2.6.1 Qiagen Therascreen EGFR RGQ PCR

The UHCW Pathology Department uses the Qiagen Therascreen EGFR RGQ PCR assay (Qiagen Ltd, Manchester, UK) for detection of *EGFR* mutations in tumour tissue samples. This assay detects a total of 29 mutations across *EGFR* exons 19 to 21, see figure 2.2.6.1. The Therascreen assay requires the preparation of eight master mixes (seven mutation detection assays and one control assay), giving a total DNA requirement of 80 ng. The assays were run in a Qiagen RotorGene QPCR instrument according to the manufacturer's instructions, and analysed according to manufacturer's instructions.

Table 2.2.6.1 Mutations tested by Qiagen Therascreen EGFR RGQ PCR Kit, information from Kit Handbook.

Qiagen Therascreen EGFR RGQ PCR Kit Mutations detectable against a background of wild-type genomic DNA
19 Deletions in exon 19 (detects mutations but does no distinguish between them)
T790M
L858R
L861Q
G719X (detects G719S, G719A and G719C but does not distinguish between them)
S768I
3 insertions in exon 20 (detects mutations but does no distinguish between them)

2.2.6.2 Life Technologies *KRAS* castPCR™

The UHCW Pathology Department uses an in house *KRAS* castPCR assay based around a custom plate design, validated and performed as recently reported (Bolton *et al.*, 2015),

using standard operating procedures. 50 ng of gDNA is required per reaction and up to 6 samples can be run per plate. The castPCR assay format is a predesigned 96-well plate in which the wells are preloaded with the primers and probes for mutation and control assays (16 in total), therefore the only reagents that need to be prepared are PCR master mix and gDNA from each of the samples to be analysed. Every sample and control must be analysed by all 16 assays (20 µl per reaction), and therefore the PCR master mix and gDNA are combined and diluted with nuclease free dH_2O to achieve the correct concentration and volume. The plates run in a Life Technologies ViiA7 instrument (Thermo Fisher Scientific) according to manufacturer's instructions. The mutations that are detectable by the *KRAS* castPCR assay are shown in figure 2.2.6.2.

Assay name	COSMIC ID	Gene	Nucleotide mutation	Amino acid change
BRAF_476_mu	476	BRAF	c.1799T>A	p.V600E
KRAS_516_mu	516	KRAS	c.34G>T	p.G12C
KRAS_517_mu	517	KRAS	c.34G>A	p.G12S
KRAS_518_mu	518	KRAS	c.34G>C	p.G12R
KRAS_520_mu	520	KRAS	c.35G>T	p.G12V
KRAS_521_mu	521	KRAS	c.35G>A	p.G12D
KRAS_522_mu	522	KRAS	c.35G>C	p.G12A
KRAS_528_mu	528	KRAS	c.37G>A	p.G13S
KRAS_529_mu	529	KRAS	c.37G>C	p.G13R
KRAS_532_mu	532	KRAS	c.38G>A	p.G13D
KRAS_552_mu	552	KRAS	c.182A>G	p.Q61R
KRAS_553_mu	553	KRAS	c.182A>T	p.Q61L
KRAS_554_mu	554	KRAS	c.183A>C	p.Q61H
KRAS_555_mu	555	KRAS	c.183A>T	p.Q61H

Figure 2.2.6.2: Exert from Life Technologies *BRAF KRAS* castPCR Kit Handbook showing mutant targets included in the assay.

Surplus DNA extracts from diagnostic samples which were tested using either the Qiagen Therascreen *EGFR* assay or the Life Technologies *KRAS* in-house castPCR assay were selected for parallel testing with the prototype GeneFirst *EGFR* and *KRAS* MMD PCR assays.

2.2.7 Quantification of DNA by the Qubit platform

UHCW Pathology Department SOPs include DNA quantification steps using the Nanodrop spectrophotometer instrument, which is standard method used in the department. For Nanodrop protocol see chapter four methods (section 4.2.4). Extracted DNA samples were analysed immediately by BMS staff after extraction where possible. If not possible they were stored at -20˚C until ready to be quantified.

As earlier quantification data (from Nanodrop) for the diagnostic DNA extracts was typically not straightforward to recover (as it did not constitute part of the diagnostic report), all samples were re-quantified using the Qubit 2.0 Nucleic Acid Analyser platform and the Qubit High Sensitivity dsDNA kit. The Qubit High Sensitivity dsDNA reagent was diluted 1/200 in Qubit High Sensitivity dsDNA buffer to create the working buffer. 199 µl of Qubit working buffer was added to 1 µl of an extracted DNA sample in a 0.5 ml optical Qubit tube. After brief mixing and centrifuging the tube (containing the working buffer and sample) was loaded into the Qubit instrument and measured according to manufacturer's instructions. The DNA quantification measurement was given in ng/ µl.

Figure 2.2.7 Qubit 2.0 Nucleic Acid Analyser Instrument.

2.2.8 GeneFirst PCR

Extracted DNA samples (from multiple sources, as described above) were analysed for mutations in *EGFR* and *KRAS* using the novel GeneFirst EGFR and KRAS mutation detection assays.

2.2.8.1 GeneFirst EGFR Kit

The assay kit was composed of an enzyme master mix and a series of primer-probe mixtures for the detection of specific mutations. The range of mutations that the assay could detect expanded over the course of the various assay versions produced by GeneFirst. A table showing the assay versions and the mutations they were designed to detect is shown below.

Table 2.2.8.1.1: Table showing the range of *EGFR* mutations detected by versions 1- 5 of the GeneFirst assay.

EGFR MASTER MIX	Target Mutations Version 1	Version 2	Version 3	Version 4	Version 5
A	19 deletions in exon 19	T790M (exon 20)	T790M (exon 20)	T790M (exon 20)	T790M (exon 20)
	T790M, L858R	L861Q (exon 21)	L861Q (exon 21)	L861Q (exon 21)	G719C (exon 18)
B	n/a	G719C (exon 18)	G719C (exon 18)	G719C (exon 18)	L861Q (exon 21)
	n/a	G719S (exon 18)	G719S (exon 18)	G719S (exon 18)	G719S (exon 18)
	n/a	G719A (exon 18)	G719A (exon 18)	G719A (exon 18)	G719A (exon 18)
C	n/a	2307_2308ins9 (exon 20)	2307_2308ins9 (exon 20)	2307_2308ins9 (exon 20)	2307_2308ins9 (exon 20)
	n/a	2310_2311insGGT (exon 20)	2310_2311insGGT (exon 20)	2310_2311insGGT (exon 20)	2310_2311insGGT (exon 20)
	n/a	2319_2320ins CAC (exon 20)	2319_2320ins CAC (exon 20)	2319_2320ins CAC (exon 20)	2319_2320ins CAC (exon 20)
	n/a	S768I (exon 20)	S768I (exon 20)	S768I (exon 20)	S768I (exon 20)
D	n/a	L858R (exon 21)	L858R (exon 21)	L858R (exon 21)	L858R (exon 21)
	n/a	19 exon 19 deletions	19 exon 19 deletions	19 exon 19 deletions	19 exon 19 deletions

Amplification reactions were set up, typically comprising of 10 µl enzyme mix, 8 µl primer-probe mix and 2 µl gDNA (1-50ng) or 2 µl dH_2O (for no template control), total reaction volume 20 µl. The exact composition of the reactions varied as the kits were updated by GeneFirst. Table 2.2.8.1.2 below shows the reaction compositions over the course of version 1 to 5 of the EGFR kit.

Table 2.2.8.1.2: Reaction compositions over all five versions of the GeneFirst EGFR PCR kit.

	EGFR Assay Version:				
	1	2	3	4	5
REAGENTS (µl)					
Taq Mix	10	10	10	10	8
EGFR Mix (A, B, C, D)	8	8	8	6	6
dH_2O	-	-	-	to 20 µl	to 20 µl
gDNA	2	2	2	1-4	1-6

PCR Reactions were run according to manufacturer's instructions using the Life Technologies ViiA7 Real Time PCR platform in 96 well plates (Microamp® Optical 96 well plates). Cycling conditions also varied as the kit was updated; table 2.2.8.1.3 shows the cycling conditions for version 1 to 5.

Table 2.2.8.1.3: Cycling conditions for versions 1- 5 of the GeneFirst EGFR PCR kit.

GENEFIRST EGFR ASSAY				
Assay Version	**Cycles**	**Temperature**	**Duration of Cycles**	**Data Collection**
1	1	95°C	9 min	
	10	95°C	6 sec	
		52°C	30 sec	
		72°C	15 sec	
	40	93°C	6 sec	
		56°C	35 sec	FAM, VIC, CY5, ROX
		60°C	30 sec	
2	1	50°C	2 min	
		95°C	9 min	
	10	95°C	4 sec	
		50°C	30 sec	
		72°C	20 sec	
	50	93°C	6 sec	
		54°C	30 sec	FAM, VIC, ROX
		60°C	30 sec	
3	1	50°C	2 min	
		95°C	8 min	
	12	95°C	6 sec	
		62°C	45 sec	
	9	90°C	6 sec	
		52°C	30 sec	
		63°C	15 sec	
	40	90°C	6 sec	
		56°C	35 sec	FAM, VIC, ROX
		60°C	15 sec	
4	1	95°C	3 min	
	10	95°C	4 sec	
		52°C	25 sec	
		72°C	15 sec	
	45	91°C	5 sec	
		56°C	35 sec	FAM, VIC, ROX
		60°C	20 sec	
5	1	50°C	2 min	
		95°C	3 min	
	10	95°C	4 sec	
		63°C	30 sec	
	10	95°C	4 sec	
		52°C	30 sec	
		72°C	20 sec	
	40	93°C	6 sec	
		56°C	30 sec	FAM, VIC, ROX
		60°C	30 sec	

2.2.8.2 GeneFirst KRAS Kit

The assay kit was composed of an enzyme master mix and a series of primer-probe mixtures for the detection of specific mutations. The range of mutations that the assay could detect expanded over the course of the various assay version produced by GeneFirst. A table showing the assay versions and the mutations they were designed to detect is shown below.

Table 2.2.8.2.1: Table showing the range of *KRAS* mutations detected by versions 1- 5 of the GeneFirst assay.

KRAS MASTER MIX	Target Mutations				
	Version 1	Version 2	Version 3	Version 4	Version 5
A	Gly12Ala	Gly12Arg	Gly12Arg	Gly12Arg	Gly12Arg
	Gly12Asp	Gly12Cys	Gly12Cys	Gly12Cys	Gly12Cys
	Gly13Asp	Gly12Val	Gly12Val	Gly12Val	Gly12Val
	Gly12Arg	-	-	-	Gly12Ser
	-	-	-	-	Gly12Asp
B	Gly12Cys	Gly12Ser	Gly12Ser	Gly12Ser	Gly12Ser
	Gly12Ser	Gly12Ala	Gly12Ala	Gly12Ala	Gly12Ala
	Gly12Val	Gly12Asp	Gly12Asp	Gly12Asp	Gly12Asp
	-	Gly13Asp	Gly13Asp	Gly13Asp	Gly13Asp

Amplification reactions were set up, typically comprising of 12.5 µl enzyme mix, 8.5 µl primer-probe mix and 2 µl gDNA (1-50ng) and 2µl dH_2O (4 µl for no template control), total reaction volume 25 µl. Reactions were thermo-cycled according to manufacturer's instructions using the Life Technologies ViiA7 Real Time PCR platform. Tables 5 and 6 below show the reaction mix compositions and modifications to the cycling conditions for version 1 to 3 of the KRAS kit.

Table 2.2.8.2.2: Reaction composition for version 1-5 of the GeneFirst *KRAS* PCR Kit.

	KRAS Assay Version:				
	1	**2**	**3**	**4**	**5**
REAGENTS (µl)					
Taq Mix	10	12.5	10	10	8
KRAS Mix (A, B)	5	8.5	8	8	6
dH_2O	0-5	3	0	0	to 20 µl
gDNA	1-5	2	2	2	1-6

Table 2.2.8.2.3: Cycling conditions for versions 1- 5 of the GeneFirst KRAS PCR kit.

GENEFIRST KRAS ASSAY				
Assay Version	**Cycles**	**Temperature**	**Duration of Cycles**	**Data Collection**
1	1	95°C	9 min	
	10	95°C	9 sec	
		51°C	30 sec	
		72°C	15 sec	
	40	93°C	6 sec	
		55°C	35 sec	FAM, VIC, ROX
		60°C	30 sec	
2	1	50°C	2 min	
		95°C	3 min	
	10	95°C	4 sec	
		50°C	30 sec	
		72°C	20 sec	
	50	93°C	6 sec	
		54°C	30 sec	FAM, VIC, ROX
		60°C	30 sec	
3	1	50°C	2 min	
		95°C	8 min	
	12	95°C	6 sec	
		62°C	45 sec	
	9	90°C	6 sec	
		52°C	30 sec	
		63°C	15 sec	
	40	90°C	6 sec	
		56°C	35 sec	FAM, VIC, ROX
		60°C	15 sec	
4	1	50°C	2 min	
		95°C	8 min	
	12	95°C	6 sec	
		62°C	45 sec	
	9	90°C	6 sec	
		52°C	30 sec	
		63°C	15 sec	
	40	90°C	6 sec	
		56°C	35 sec	FAM, VIC, ROX
		60°C	15 sec	
5	1	50°C	2 min	
		95°C	3 min	
	10	95°C	4 sec	
		63°C	30 sec	
	10	95°C	4 sec	
		52°C	30 sec	
		72°C	20 sec	
	40	93°C	6 sec	
		56°C	30 sec	FAM, VIC, ROX
		60°C	30 sec	

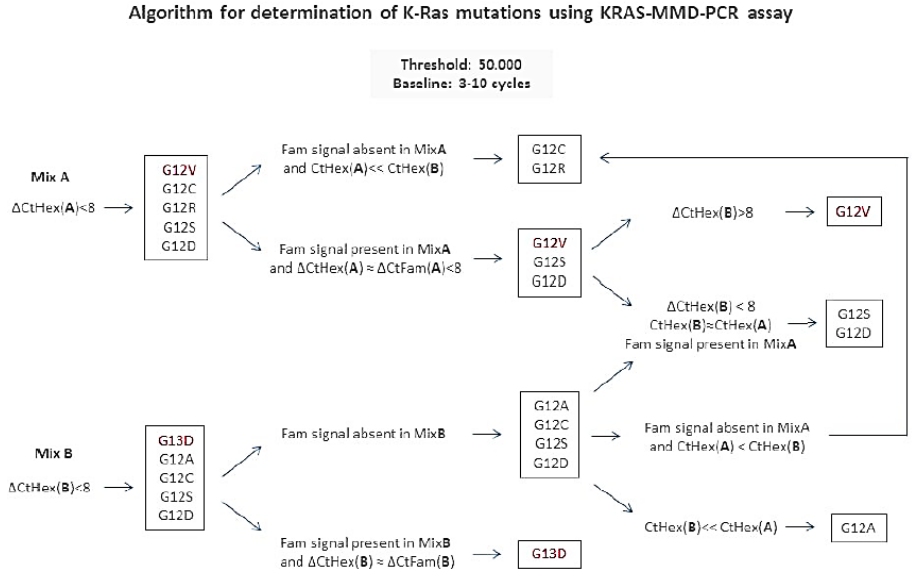

Figure 2.2.8.2: Additional analysis algorithm for interpretation of GF *KRAS* PCR version 5 results. (From GeneFirst kit insert)

2.2.8.3 Next Generation Sequencing of Tissue DNA

Libraries were prepared from extracted clinical samples using the Life Technologies Ion Torrent AmpliSeq™ 200 library kit, according to the manufacturer's guidelines. Just 10ng of template gDNA was used per sample for each of the library preparations. Target regions were amplified using the European OncoNetwork Colon and Lung Cancer 22-gene panel which is commercially available as the Ion AmpliSeq™ Colon and Lung Panel (ThermoFisher). Figure 2.2.8.3.1 below gives an overview of the Colon and Lung Panel primer pool.

Table 2.2.8.3.1: Specifications of European OncoNetwork 22-gene panel, included in Ion AmpliSeq™ Colon and Lung Cancer Panel (from Thermo Fisher OncoNetwork panel booklet). This NGS panel was later included in the CE-IVD Oncomine Solid Tumour DNA kit for NGS library preparation.

Ion AmpliSeq™ Colon and Lung Panel	
Targets	Hotspots and targeted regions (totaling 14.6 kb) in 22 genes implicated in colon and lung cancer
Genes	*KRAS, EGFR, BRAF, PIK3CA, AKT1, ERBB2, PTEN, NRAS, STK11, MAP2K1, ALK, DDR2, CTNNB1, MET, TP53, SMAD4, FBXW7, FGFR3, NOTCH1, ERBB4, FGFR1, FGFR2*
Amplicon length	90 amplicons with an average length of 162 bp (ideal for FFPE samples)
Primer pool	90 pairs of primers in a single pool
Input DNA required	10 ng
Recommended multiplexing	8 samples per Ion 316™ Chip with at least 500x sequencing coverage

The amplified products from each sample were ligated with separate Ion Express barcode sequences to enable simple differentiation of the samples by the Ion Torrent platform. The libraries were then purified and quantified using The Ion AmpliSeq™ Q-PCR method. Quantified libraries were then combined into solutions of equal concentration (three samples/ libraries at a time). These were then further processed and loaded onto the Ion PGM 314 chips using the Ion Chef platform. The loaded chips were then sequenced using the Ion Torrent PGM Instrument. Sequence analysis parameters were set up in advance and performed automatically by the Ion Torrent Server running Variant Caller software. Data from each library provided a comprehensive output of all mutations detected, included known and novel.

Sequencing was performed as previously described (Tops *et al.*, 2015), with the following changes to accommodate Ion Chef loading of the Ion Torrent 316 or 314 chips. Stored or newly extracted DNA was checked for content and quality using a Qubit 2.0 Fluorometer. 10 ng of gDNA from each of the samples chosen for NGS analysis was combined with the Ampliseq™ reagents and primer pool for the OncoNetwork 22 gene panel (Tops *et al.*, 2015)

and amplified for 22 cycles. After initial amplification the amplified products are partially digested before Ion Express Adapters and Barcode sequences are ligated to the library fragments. Following barcoding the libraries are cleaned up using a magnetic bead method. The cleaned up products are then quantified by the AmpliSeq™ Q-PCR method. Once the libraries have been successfully quantified they are combined and diluted to 50 pM. For Ion Torrent 314 chips, 3 libraries were combined per chip. These library pools were then loaded into the Ion Chef instrument for further library preparation and chip loading. The loaded 314 chips were then run on the Ion Torrent PGM instrument according to manufacturer's instructions. The Variant Caller plugin (including in the provided Ion Suite software) was used to analyse the aligned sequence data for the identification of hotspot mutations and novel variants.

Figure 2.2.8.3.2 (Left) Life Technologies Ion Chef Instrument. (Right) Ion Torrent Personal Genome Machine (PGM).

2.3 RESULTS

Over the course of the project, the GeneFirst EGFR and KRAS PNA/ LNA PCR assay went through several developmental versions (five for both the EGFR and KRAS kits), based on data and feedback from the project. All versions of the assays were tested from October 2013 to December 2014 using pre-extracted nucleic acid samples obtained from the Pathology Department, UHCW. The samples were all FFPE diagnostic tissue samples from lung and colorectal cancer patients, unless otherwise stated. Tables 2.8.1.2, 2.8.1.3, 2.8.2.2 and 2.8.2.3 in the Methods section shows how the PCR protocol and cycling conditions changed with each new version of the assay.

GENEFIRST EGFR ASSAY

A total of 142 tissue sample analyses were performed using various versions of the GeneFirst EGFR PCR assay. Since the beginning of the project, the EGFR assay progressed through five successive versions. All samples used to test the GeneFirst EGFR Assay were DNA samples extracted from tissue.

2.3.1 GF EGFR Assay Version 1

The first version of the assay provided single tube amplification and detection of three significant *EGFR* mutations, using three different fluorescent dyes (FAM, VIC and Cy5). The assay also featured an endogenous control for monitor reaction efficiency, which used the ROX dye reporter.

Table 2.3.1.1 show a summary of the samples tested using the EGFR version 1 assay, and the percentage concordance between the currently used Qiagen Therascreen EGFR assay and the prototype GeneFirst EGFR assay. The percentage concordance for mutant positive samples and wild-type (WT) samples was 68.8% and 60% respectively.

Table 2.3.1.1: Summary table showing the concordance % and discrepancy % of the GeneFirst EGFR PCR assay version 1.

		GeneFirst EGFR v1			**Sample Concordance %**	
		Match	Discrepant	Totals	Mutant Samples	WT samples
Therascreen	Mutant	11	5	16	68.8	
	WT	3	2	5		60.0
	Totals	14	7	21		
Total	**Concordance %**	66.7				
	Discrepancy %		33.3			

Whilst a majority (66.7%) of sample results agreed with the original Therascreen results, overall the level of discrepancy was 33.3%, a significant level, and not suitable for diagnostic use. The most frequent discrepancy encountered was the occurrence of false positives, where mutant samples would give positive results for unexpected mutations, or where WT samples would give a positive result. For more results of individual samples run, see chapter two appendix, table A2.1.1.

Table 2.3.1.2 Endogenous control results from samples tested with EGFR assay version 1. The Mutation Type field refers to the results determined by previous testing with the Qiagen Therascreen assay. ROX = ROX reporter signal, NTC = No template control, WT = Wild-type.

			ROX	PCR VALIDATION		
Study ID	**Mutation type**	**DNA (ng, total)**	**Ct Mean**	**Control Pos/ Neg**	**Assay Pass/ Fail**	**Comments**
EGFRv1 01	Exon 19 Del	7.8	30.189	POS	PASS	
EGFRv1 02	Exon 19 Del	16.0	36.170	POS	PASS	
EGFRv1 03	Exon 19 Del	16.8		NEG	FAIL	Endogenous control fail
EGFRv1 04	Exon 19 Del	11.2		NEG	FAIL	Endogenous control fail
NTC		0.0		NEG	PASS	
EGFRv1 05	L858R	25.6	32.480	POS	PASS	
EGFRv1 06	L858R	15.8		NEG	FAIL	Endogenous control fail
EGFRv1 07	L858R	16.8		NEG	FAIL	Endogenous control fail
NTC		0.0		NEG	PASS	
EGFRv1 08	T790M	33.4	23.994	POS	PASS	
EGFRv1 09	T790M	18.4	29.418	POS	PASS	
EGFRv1 10	WT	12.0	32.495	POS	PASS	
EGFRv1 11	WT	9.0	28.436	POS	PASS	
EGFRv1 12	WT	25.2	29.318	POS	PASS	
EGFRv1 13	WT	6.0	29.114	POS	PASS	
EGFRv1 15	T790M + L858R	16.8	31.982	POS	PASS	
EGFRv1 16	T790M	34.0	29.212	POS	PASS	
EGFRv1 18	T790M	30.0	30.740	POS	PASS	
NTC				NEG	PASS	
EGFRv1 17	Exon 19 Del	52.6	30.328	POS	PASS	
EGFRv1 20	Exon 19 Del	2.5	30.803	POS	PASS	
NTC				NEG	PASS	
EGFRv1 19	L861Q	1.6	35.811	POS	PASS	
EGFRv1 21	S768I	13.5		NEG	FAIL	Endogenous control fail
EGFRv1 14	WT	34.6	30.163	POS	PASS	
				PCR Pass:	20	
				PCR Fail:	5	
				Failure rate:	**20.0%**	

Table 2.3.1.2 shows the performance of the endogenous control assay (ROX fluorophore) in GF EGFR assay version 1. Of 25 samples, five have given incorrect results, a failure rate of 20%. All NTC (no template controls) gave a negative signal.

2.3.2 GF EGFR Assay Version 2

The second version of the assay expanded the assay's range dramatically. By using four separate master mixes, the assay was now able to identify more mutations, see Methods table 2.8.1.1. This also meant the assay was more complicated to set up, had a much higher volume requirement from the samples and reduced the number of samples that could be analysed per 96 well plate. As each sample had to be run in duplicate over four master mixes, therefore each sample took up eight wells in the plate, and needed sufficient DNA for all reactions to occur optimally. This also posed a significant problem in terms of sourcing adequate volumes of sample material. The samples obtained from the Pathology were all previously extracted and tested, resulting in the available volumes varying greatly. The average remaining volume of the previously tested Pathology lab samples was between 15-20 µl, although there was often far less than this remaining. As the EGFR assay required a total 16 µl per sample, the limited volumes available made it sometimes impossible to fully test many samples.

Table 2.3.2.1 shows the percentage concordance between the currently used Qiagen Therascreen EGFR assay and the prototype GeneFirst EGFR assay. The percentage concordance for mutant positive samples and wild-type samples was 46.2% and 74.1% respectively. A table of the individual samples tested using version of the GeneFirst EGFR assay is included in the chapter 2 appendix (table A2.1.2).

Table 2.3.2.1: Summary table showing the concordance % and discrepancy % of the GeneFirst EGFR PCR assay version 2. WT= Wild-type.

		GeneFirst EGFR v2			Sample Concordance %	
		Match	Discrepant	Totals	Mutant Samples	WT samples
Therascreen	Mutant	12	14	26	46.2	
	WT	20	7	27		74.1
	Totals	32	21	53		
Total	**Concordance %**	60.4				
	Discrepancy %		39.6			

Overall the more complicated version 2 of the assay achieved a lower % concordance than the previous version, 60.4% for version 2 compared to 66.6% in version 1.

Table 2.3.2.2: Limit Of Detection (LOD) experiments using mutant samples diluted to varying degrees with WT samples. The Mutation Type field refers to the results determined by previous testing with the Qiagen Therascreen assay. For each reaction mix: green +/- symbols indicate true results, red +/- symbols indicate false results.

Study ID	Mutation type	Dilution	DNA (ng)	Result (Reaction Mix A)	Result (Reaction Mix D)	Limit of Detection
EGFRv2 48	L858R	1	15	+	+	**0.5 ng**
		1/10	1.5	+	+	
		1/20	1	+	+	
		1/50	0.5	-	+	
EGFRv2 52	T790M	1	30	-	-	**1.5 ng**
		1/10	3	+	-	
		1/20	1.5	+	-	
		1/50	0.6	-	-	

In addition to the standard mutation testing, Limit of Detection (LoD) experiments were carried out, see table 2.3.2.2. Here a positive sample was diluted using WT samples. These diluted samples were run on the assay to assess the Limit of Detection. The version 2 EGFR assay achieved a 0.5 ng LoD for sample 48, and 1.5 ng for sample 52, showing a good level of sensitivity. However, the assay also produced a number of false positives in one of master mixes for sample 48 (master mix A), and for sample 52 the undiluted sample was not detected.

To summarise, version 2 of the EGFR assay showed a promising LoD sensitivity, but still a poor specificity owing to a high level of false positives.

Table 2.3.2.4 Endogenous control results from samples tested with EGFR assay version 2. The Mutation Type field refers to the results determined by previous testing with the Qiagen Therascreen assay. NTC = No template control, WT = Wild-type.

Study ID	Mutation type	DNA (ng, total)	PCR VALIDATION Controls Pos/ Neg	PCR VALIDATION Assay Pass/ Fail	Study ID	Mutation type	DNA (ng, total)	PCR VALIDATION Controls Pos/ Neg	PCR VALIDATION Assay Pass/ Fail
EGFRv2 10	T790M	18.40	POS	PASS	EGFRv2 31	WT	0.52	POS	PASS
EGFRv2 14	T790M	16.40	POS	PASS	EGFRv2 32	WT	73.92	POS	PASS
EGFRv2 12	L861Q	25.20	POS	PASS	EGFRv2 33	WT	22.56	POS	PASS
EGFRv2 09	WT	25.60	POS	PASS	EGFRv2 34	WT	4.14	POS	PASS
EGFRv2 51	L858R	16.80	POS	PASS	EGFRv2 35	WT	9.86	POS	PASS
EGFRv2 16	L858R	9.00	POS	PASS	NTC			NEG	PASS
EGFRv2 08	WT	7.80	POS	PASS	EGFRv2 36	WT	0.36	POS	PASS
EGFRv2 01	Exon 19 Del	16.00	POS	PASS	EGFRv2 37	Exon 19 Del	0.43	POS	PASS
EGFRv2 02	Exon 19 Del	16.80	POS	PASS	EGFRv2 38	Exon 19 Del	1.64	POS	PASS
EGFRv2 03	Exon 19 Del	16.80	POS	PASS	EGFRv2 39	G719X	1.06	POS	PASS
NTC			NEG	PASS	NTC			NEG	PASS
EGFRv2 13	T790M	80.00	POS	PASS	EGFRv2 40	WT	2.34	POS	PASS
EGFRv2 11	T790M	80.00	POS	PASS	EGFRv2 41	WT	17.68	POS	PASS
EGFRv2 06	T790M	18.40	POS	PASS	EGFRv2 42	WT	1.39	POS	PASS
EGFRv2 07	T790M	18.40	POS	PASS	EGFRv2 43	WT	37.40	POS	PASS
EGFRv2 50	L858R	25.60	POS	PASS	EGFRv2 44	WT	25.90	POS	PASS
EGFRv2 05	L858R	16.80	POS	PASS	EGFRv2 45	WT	23.00	POS	PASS
NTC			NEG	PASS	EGFRv2 47	WT	31.40	POS	PASS
EGFRv2 17	Exon 19 Del	56.96	POS	PASS	NTC			NEG	PASS
EGFRv2 18	S768I	13.80	POS	PASS	EGFRv2 46	WT	30.00	POS	PASS
EGFRv2 19	WT	34.60	POS	PASS	NTC			NEG	PASS
EGFRv2 20	L858R	15.80	POS	PASS	EGFRv2 49	L858R	2.80	POS	PASS
EGFRv2 21	WT	12.00	POS	PASS	EGFRv2 04	L858R	1.50	POS	PASS
EGFRv2 22	WT	0.48	POS	PASS	EGFRv2 48	L858R	1.00	POS	PASS
EGFRv2 23	WT	0.40	POS	PASS	EGFRv2 15	L858R	0.50	POS	PASS
EGFRv2 24	WT	0.40	POS	PASS	NTC			NEG	PASS
EGFRv2 25	WT	15.32	POS	PASS	EGFRv2 55	T790M	30.00	POS	PASS
EGFRv2 26	WT	0.32	POS	PASS	EGFRv2 52	T790M	3.00	POS	PASS
EGFRv2 27	WT	18.00	POS	PASS	EGFRv2 53	T790M	1.50	POS	PASS
NTC			NEG	PASS	EGFRv2 54	T790M	0.60	POS	PASS
EGFRv2 28	WT	2.90	POS	PASS				**PCR Pass:**	63
EGFRv2 29	WT	39.20	POS	PASS				**PCR Fail:**	0
EGFRv2 30	WT	2.88	POS	PASS				**Failure rate:**	**0.0%**

Table 2.3.2.4 shows the performance of the ROX endogenous control assay for all individual samples tested with the GF EGFR assay version 2. All samples gave the expected results (samples ROX positive, NTC ROX negative).

2.3.3 GF EGFR Assay Version 3

Version 3 of the EGFR assay retained the same expanded panel of mutations and four master mix format. Where possible, samples that had been previously tested on version 2 were re-tested on version 3, although this was not possible in a number of cases due to the very low volume of original sample. There was also a severely limited input of new samples, due to the generally low occurrence of *EGFR* positive samples processed by the Pathology Department. There was also a period where the Pathology Department temporarily suspended its *EGFR* testing service, which resulted in a complete absence of new samples during this period.

Table 2.3.3.1: Summary table showing the concordance % and discrepancy % of the GeneFirst EGFR PCR assay version 3. WT= Wild-type.

		GeneFirst EGFR v3			**Sample Concordance %**	
		Match	Discrepant	Totals	Mutant Samples	WT samples
Therascreen	Mutant	5	11	16	31.3	
	WT	4	3	7		57.1
	Totals	9	14	23		
Total	**Concordance %**	39.1				
	Discrepancy %		60.9			

Table 2.3.3.1 shows the levels of concordance and discrepancy in version 3 of the assay to be 39.1% and 60.9% respectively. Compared to version 2 this represents a 21.3% reduction in concordance with the original Therascreen results. For the individual concordant and

discrepant sample results tested by version 3 of the EGFR assay, see chapter two appendix, table A2.1.3.

In addition to the low concordance with Therascreen, another problem with the GeneFirst EGFR assay appeared in version 3. In many analysis runs, the No Template Control (NTC) negative control samples started giving a positive signal for the endogenous control (ROX channel). This was initially considered to be accidental user contamination; however successive repeats and reactions prepared with only master mix (i.e. reagents only, no DNA or nuclease free water) continued to produce this endogenous control positive signal. Examples of these experiments are shown in chapter two appendix, section A2.1 table A2.1.6 and figure A2.1.

To further investigate this phenomenon the PCR products from the no template control experiments were run on an agarose gel and examined under UV to assess the presence of primer dimers or additional amplicons. Figure 2.3.3 shows the gel image from this analysis, there is no evidence of primer dimers or amplicons, suggesting that the signals detected in the ROX channel may be due to technical reasons associated with the ROX probe.

Figure 2.3.3 Image of GF EGFR version 3 assay PCR products run on a 2% agarose gel, visualised under UV. The loading wells are located towards the top of the image, above to the 1000 bp ladder position. There is no evidence of primer dimers or amplicons on the gel.

Table 2.3.3.2 Endogenous control results from samples tested with EGFR assay version 3. Ct values from failed assays are highlighted in red. Cells containing anomalous Ct values from reactions where no template was loaded are given a black border. ROX A = ROX reporter signal in master mix A, ROX B = ROX reporter signal in master mix B, ROX C = ROX reporter signal in master mix C, ROX D = ROX reporter signal in master mix D, NTC = No template control, WT = Wild-type.

			ROX A		ROX B		ROX C		ROX D		
Study ID	Mutation type	DNA (ng, total)	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Assay Pass/ Fail
EGFRv3 01	T790M	60.0	24.447	POS	28.650	POS	26.799	POS	33.274	POS	FAIL
EGFRv3 02	T790M	36.8	20.186	POS	22.001	POS	19.317	POS	20.538	POS	FAIL
EGFRv3 03	T790M	18.4	23.493	POS	28.560	POS	27.294	POS	27.982	POS	FAIL
EGFRv3 04	L858R	5.6	21.690	POS	22.663	POS	20.723	POS	25.888	POS	PASS
EGFRv3 05	Exon 19 Del	113.9	24.932	POS	25.520	POS	23.605	POS	24.894	POS	PASS
EGFRv3 07	T790M	78.5	23.036	POS	23.281	POS	21.714	POS	22.860	POS	PASS
EGFRv3 06	T790M	26.2	23.148	POS	23.892	POS	21.595	POS	23.401	POS	PASS
EGFRv3 10	WT	10.0	29.009	POS	29.328	POS	26.388	POS	28.136	POS	PASS
EGFRv3 08	WT	149.6	27.130	POS	27.096	POS	24.678	POS	24.805	POS	PASS
EGFRv3 09	WT	74.8	24.813	POS	24.921	POS	23.063	POS	24.461	POS	PASS
NTC	-	0.0	34.240	POS	30.036	POS	27.992	POS	29.199	POS	FAIL
EGFRv3 11	WT	207.2	35.255	POS	35.672	POS	35.183	POS	38.240	POS	PASS
EGFRv3 12	WT	51.8	26.392	POS	27.146	POS	23.619	POS	27.247	POS	FAIL
EGFRv3 13	Exon 19 Del	0.9	27.652	POS	28.320	POS	30.154	POS	27.278	POS	PASS
EGFRv3 14	WT	69.2	27.787	POS	28.284	POS	30.979	POS	29.106	POS	PASS
EGFRv3 15	L858R	16.8	28.108	POS	29.497	POS	27.319	POS	27.026	POS	PASS
EGFRv3 16	WT	0.6	28.226	POS	29.567	POS	26.735	POS	29.208	POS	FAIL
EGFRv3 22	Exon 19 Del	1.7	24.157	POS	23.661	POS	23.181	POS	25.761	POS	PASS
EGFRv3 23	L861Q	60.8	24.095	POS	22.743	POS	21.010	POS	22.895	POS	PASS
EGFRv3 24	L858R	16.2	25.160	POS	25.153	POS	23.504	POS	35.112	POS	PASS
EGFRv3 25	L861Q	0.9	29.923	POS	29.469	POS	32.514	POS	29.910	POS	PASS
EGFRv3 26	Exon 19 Del	18.2	24.120	POS	24.460	POS	22.697	POS	27.985	POS	PASS
NTC	-	0.0	31.221	POS	30.389	POS	28.150	POS		POS	FAIL
NTC	-	0.0	29.802	POS		NEG	31.689	POS	29.680	POS	FAIL
EGFRv3 27	L861Q	60.8	24.113	POS	22.617	POS	27.248	POS	27.705	POS	PASS
EGFRv3 28	L861Q	0.9	31.556	POS	29.056	POS	28.673	POS	29.838	POS	PASS
NTC	-	0.0	30.716	POS	29.824	POS	28.719	POS	32.390	POS	FAIL
										PCR Pass:	18
										PCR Fail:	9
										Failure rate:	33.3%

Table 2.3.3.2 shows the results from the ROX endogenous control assay for all samples tested with GF EGFR assay version 3. 9/18 samples failed the control assay in at least one master mix. The strangest results are those where no template was added (cells given a black border), as these would be expected to be negative.

2.3.4 GF EGFR Assay Version 4

Version 4 of the EGFR assay saw a minor improvement in concordance with the Therascreen results. Overall concordance was 46.67% which is a higher level of concordance than was achieved by version 3 (39.12%), however less samples were available for testing in version 4 (23 samples for version 3, 15 samples for version 4). The results for the individual samples tested can be found in table A2.1.4 (chapter two appendix).

Table 2.3.4.1 shows a breakdown of concordance across the mutant and wild-type samples tested. The assay gave 100% concordance with the wild-type samples tested, however only two samples were tested. The recurring problem of false positives/ discrepant results with mutant samples from previous versions of the assay were also present in this version: 8 of 13 mutant samples gave discrepant results compared to the original Therascreen results.

Table 2.3.4.1: Summary table showing the concordance % and discrepancy % of the GeneFirst EGFR PCR assay version 4. WT = Wild-type.

		GeneFirst EGFR v4			Sample Concordance %	
		Match	Discrepant	Totals	Mutant Samples	WT samples
Therascreen	Mutant	5	8	13	38.5	
	WT	2	0	2		100.0
	Totals	7	8	15		
Total	Concordance %	46.7				
	Discrepancy %		53.3			

Table 2.3.4.2 Endogenous control results from samples tested using EGFR assay version 4. Anomalous Ct values are highlighted in red. ROX A = ROX reporter signal in master mix A, ROX B = ROX reporter signal in master mix B, ROX C = ROX reporter signal in master mix C, ROX D = ROX reporter signal in master mix D, NTC = No template control, WT = Wild-type.

			ROX A		ROX B		ROX C		ROX D		
Study ID	Mutation type	DNA (ng, total)	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Assay Pass/ Fail
EGFRv4 14	Exon 19 Del	57.0	19.831	POS	19.702	POS	18.486	POS	18.643	POS	PASS
EGFRv4 02	T790M	78.5	17.395	POS	17.309	POS	18.557	POS	17.336	POS	PASS
EGFRv4 03	Exon 19 Del	1.7	22.789	POS	22.314	POS	20.506	POS	20.236	POS	PASS
EGFRv4 04	WT	35.6	21.595	POS	21.207	POS	20.252	POS	20.446	POS	PASS
EGFRv4 05	T790M	17.5	18.669	POS	18.435	POS	17.422	POS	18.061	POS	PASS
EGFRv4 06	L858R	33.6	22.201	POS	21.722	POS	20.234	POS	20.625	POS	PASS
EGFRv4 07	L858R	2.6	18.764	POS	19.072	POS	17.842	POS	18.491	POS	PASS
EGFRv4 08	L858R	1.3	19.160	POS	19.187	POS	17.898	POS	19.160	POS	PASS
EGFRv4 09	L858R	0.5	19.162	POS	18.942	POS	17.996	POS	19.144	POS	PASS
NTC		0.0	21.906	POS	21.865	POS	20.648	POS	20.559	POS	FAIL
NTC		0.0	22.430	POS	21.692	POS	20.556	POS	20.680	POS	FAIL
NTC		0.0	22.178	POS	21.766	POS	20.885	POS	20.693	POS	FAIL
EGFRv4 10	Exon 19 Del	32.2	24.345	POS	27.789	POS	25.152	POS	26.784	POS	PASS
EGFRv4 11	Exon 19 Del	11.2	24.005	POS	28.999	POS	25.936	POS	27.874	POS	PASS
EGFRv4 12	Exon 19 Del	16.8	24.010	POS	29.503	POS	26.360	POS	28.560	POS	PASS
EGFRv4 13	L858R	31.6	24.134	POS	29.316	POS	26.196	POS	28.800	POS	PASS
EGFRv4 01	Exon 19 Del	13.4	22.885	POS	25.216	POS	22.221	POS	23.619	POS	PASS
EGFRv4 15	WT	30.6	23.859	POS	27.863	POS	22.899	POS	23.564	POS	PASS
NTC		0.0	24.278	POS	32.091	POS	26.994	POS	31.514	POS	FAIL
NTC		0.0	23.602	POS	32.112	POS	26.228	POS	30.334	POS	FAIL
										PCR Pass:	15
										PCR Fail:	5
										Failure rate:	25.0%

Table 2.3.4.2 shows the results of the ROX endogenous control from the GF EGFR assay version 4. In this set of samples all the NTCs failed the ROX endogenous control assay, including two 'blank' samples (reactions prepared without addition of water as a substitute for sample i.e. taq mix and primer/ probe mix only). This strongly suggests an issue with one or both of the assays reagents. All samples gave positive signals in the endogenous control assay. Unlike the testing of the previous assay version, all samples had sufficient volume to test all master mixes, however the validity of these results cannot be assessed due to the positive signals given by the NTCs.

2.3.5 GF EGFR Assay Version 5

Version 5 was the final version of the EGFR assay made available by GeneFirst for this project. 30 samples were tested, 11 wild-types and 19 mutants. Unfortunately all the samples panel either contained very low remaining sample volume (after routine testing by the Pathology department) or were of low DNA concentration. 12 samples (11 wild-type and 1 mutant) contained DNA concentrations that were below the level which could be accurately quantified by the Qubit instrument. The results for the individual samples tested with this final version of the assay can be found in table A2.1.5 (chapter two appendix). Some samples are indicated with "Unknown (low)" in the DNA loading field of the table.

Overall concordance with the original Therascreen results for this version of the assay was 26.67%. This is a reduction of 20% compared to version 4. A breakdown of the results are shown in table 2.3.5.1. Despite this being the most recent version of the assay, the performance of the assay was poorer than the previous two versions. Again, the persistent issue of false positive/ discrepant results were a feature of the version 5 dataset. Only 2 of

17 (10.5%) mutant samples and 6 of 11 (55%) wild-type samples matched the original Therascreen results.

Table 2.3.5.1: Summary table showing the concordance % and discrepancy % of the GeneFirst EGFR PCR assay version 5. WT = Wild-type.

		GeneFirst EGFR v5			Sample Concordance %	
		Match	Discrepant	Totals	Mutant Samples	WT samples
Therascreen	Mutant	2	17	19	10.5	
	WT	6	5	11		54.5
	Totals	8	22	30		
Total	**Concordance %**	26.7				
	Discrepancy %		73.3			

To accurately assess the performance of version 5 of the EGFR assay, the data was reanalysed to examine the results given by each of the assays four separate reaction master mixes. The EGFR assay consisted of four master mixes, with each mix using two reporter fluorophores (VIC and FAM, linked to mutant target probes) therefore it was possible to generate eight signals per sample. From our dataset of 30 samples, it is therefore possible to have up to generate 240 PCR signals across the eight reactions. The results from all four master mixes are shown in table 2.3.5.2. In total there were 125 signals detected, of which 41 (32.8%) were either false positives or false negatives. Of the individual master mixes, the data suggests that master mix C performed the best, with only 5 false results, whereas master mix D performed the worst, with 19 false results.

Table 2.3.5.2: Complete analysis of PCR results from all GF EGFRv5 master mixes. Designations 'True Positive', 'True Negative', 'False Positive' and 'False Negative' are assigned based on the original Therascreen results.

Study ID	Mutation type	GF EGFRv5 PCR Master Mix				Result			
		A	B	C	D	True Pos	True Neg	False Pos	False Neg
EGFRv5 09	Exon 19 Del	FALSE POS	TRUE NEG	TRUE NEG	TRUE POS	1	2	1	0
EGFRv5 01	Exon 20 Ins	TRUE NEG	TRUE NEG	TRUE POS	FALSE POS	1	2	1	0
EGFRv5 02	L861Q and G719X	FALSE NEG	TRUE POS x2	TRUE NEG	FALSE POS	2	1	1	1
EGFRv5 03	Exon 19 Del	TRUE NEG	TRUE NEG	TRUE NEG	TRUE POS	1	3	0	0
EGFRv5 04	T790M	FALSE NEG	FALSE POS	TRUE NEG	FALSE POS	0	1	2	1
EGFRv5 05	L861Q	TRUE NEG	TRUE POS, FALSE POS	TRUE NEG	FALSE POS	1	2	2	0
EGFRv5 06	L858R	TRUE NEG	FALSE POS	FALSE POS	FALSE NEG	0	1	2	1
EGFRv5 07	T790M	FALSE NEG	FALSE POS	TRUE NEG	FALSE POS	0	1	2	1
EGFRv5 08	S768I	TRUE NEG	TRUE NEG	TRUE POS	FALSE POS	1	2	1	0
EGFRv5 18	Exon 19 Del	TRUE NEG	TRUE NEG	TRUE NEG	TRUE POS	1	3	0	0
EGFRv5 10	Exon 20 Ins	TRUE NEG	TRUE NEG	TRUE POS	FALSE POS x2	1	2	2	0
EGFRv5 11	L861Q and G719X	FALSE POS, FALSE NEG	TRUE POS x2	TRUE NEG	FALSE POS	2	1	2	1
EGFRv5 12	Exon 19 Del	TRUE NEG	FALSE POS	TRUE NEG	TRUE POS	1	2	1	0
EGFRv5 13	T790M	FALSE NEG	TRUE NEG	TRUE NEG	FALSE POS	0	2	1	1
EGFRv5 14	L861Q	TRUE NEG	TRUE POS	TRUE NEG	FALSE POS	1	2	1	0
EGFRv5 15	L858R	FALSE POS	FALSE POS	TRUE NEG	FALSE POS	0	1	3	0
EGFRv5 16	T790M	TRUE POS	TRUE NEG	TRUE NEG	FALSE POS	1	2	1	0
EGFRv5 17	S768I	TRUE NEG	TRUE NEG	TRUE POS	FALSE POS	1	2	1	0
EGFRv5 21	WT	TRUE NEG	TRUE NEG	TRUE NEG	TRUE NEG	0	4	0	0
EGFRv5 22	WT	TRUE NEG	TRUE NEG	TRUE NEG	TRUE NEG	0	4	0	0
EGFRv5 23	WT	TRUE NEG	TRUE NEG	TRUE NEG	TRUE NEG	0	4	0	0
EGFRv5 19	WT	FALSE POS	TRUE NEG	TRUE NEG	FALSE POS	0	2	2	0
EGFRv5 20	Del	TRUE NEG	TRUE NEG	TRUE NEG	FALSE NEG	0	3	0	1
EGFRv5 24	WT	TRUE NEG	TRUE NEG	TRUE NEG	TRUE NEG	0	4	0	0
EGFRv5 29	WT	TRUE NEG	TRUE NEG	TRUE NEG	TRUE NEG	0	4	0	0
EGFRv5 25	WT	TRUE NEG	FALSE POS	FALSE POS	TRUE NEG	0	2	2	0
EGFRv5 26	WT	TRUE NEG	TRUE NEG	FALSE POS	FALSE POS	0	2	2	0
EGFRv5 27	WT	TRUE NEG	FALSE POS	FALSE POS	FALSE POS	0	1	3	0
EGFRv5 28	WT	TRUE NEG	TRUE NEG	FALSE POS	TRUE NEG	0	3	1	0
EGFRv5 30	WT	TRUE NEG	TRUE NEG	TRUE NEG	TRUE NEG	0	4	0	0
Totals:	**TRUE**	**22**	**25**	**23**	**12**	**15**	**69**	**34**	**7**
	FALSE	**9**	**7**	**5**	**19**				

The data from table 2.2.5.2 was used to calculate the sensitivity and specificity of the version 5 assay. The results of these calculations are shown in table 3.2.5.3. The GF EGFR version 5 assay achieved a sensitivity and specificity of 68.2% and 67.0% respectively.

Table 2.3.5.3: Summary of GF EGFR version 5 data.

GeneFirst EGFR Assay version 5				
Test	**Mutant Present**		**Mutant absent**	
Positive	True Positive	15	False Positive	34
Negative	False Negative	7	True Negative	69

Statistic	**Value**
Sensitivity	**68.2%**
Specificity	**67.0%**
Positive Predictive Value	**30.6%**
Negative Predictive Value	**90.8%**

Table 2.3.5.4 Endogenous control results from samples tested using the EGFR version 5 assay. Samples with DNA concentrations below the lower quantifiable range were designated 'unknown'. ROX A = ROX reporter signal in master mix A, ROX B = ROX reporter signal in master mix B, ROX C = ROX reporter signal in master mix C, ROX D = ROX reporter signal in master mix D, NTC = No template control, WT = Wild-type.

			ROX A		ROX B		ROX C		ROX D		
Study ID	Mutation type	DNA (ng, total)	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Assay Pass/ Fail
EGFRv5 01	Exon 19 Del	30	23.134	POS	28.003	POS	25.352	POS	24.045	POS	PASS
EGFRv5 02	Exon 20 Ins	30	21.175	POS	25.900	POS	22.624	POS	23.324	POS	PASS
EGFRv5 03	L861Q and G719X	30		NEG	31.618	POS	26.254	POS	28.079	POS	FAIL
EGFRv5 04	Exon 19 Del	30	21.406	POS	26.556	POS	23.608	POS	25.294	POS	PASS
EGFRv5 05	T790M	30	22.427	POS	27.174	POS	22.891	POS	23.724	POS	PASS
EGFRv5 06	L861Q	30	22.427	POS	27.211	POS	23.094	POS	25.043	POS	PASS
EGFRv5 07	L858R	30		NEG		NEG	35.664	POS		NEG	FAIL
EGFRv5 08	T790M	30	22.169	POS	27.137	POS	23.467	POS	24.119	POS	PASS
EGFRv5 09	S768I	30	23.633	POS	29.029	POS	25.153	POS	25.799	POS	PASS
NTC		0		NEG		NEG		NEG		NEG	PASS
EGFRv5 10	Exon 19 Del	15	22.997	POS	27.510	POS	24.121	POS	24.274	POS	PASS
EGFRv5 11	Exon 20 Ins	15	21.364	POS	25.800	POS	22.607	POS	23.523	POS	PASS
EGFRv5 12	L861Q and G719X	15	21.782	POS	27.274	POS	23.397	POS	24.312	POS	PASS
EGFRv5 13	Exon 19 Del	15	22.673	POS	27.365	POS	24.074	POS	24.663	POS	PASS
EGFRv5 14	T790M	15	20.775	POS	25.400	POS	22.101	POS	23.096	POS	PASS
EGFRv5 15	L861Q	15	22.766	POS	27.961	POS	23.937	POS	24.853	POS	PASS
EGFRv5 16	L858R	15	26.743	POS	39.233	POS	27.878	POS	36.766	POS	PASS
EGFRv5 17	T790M	15	20.693	POS	25.101	POS	21.914	POS	22.526	POS	PASS
EGFRv5 18	S768I	15	22.614	POS	27.341	POS	23.800	POS	24.778	POS	PASS
NTC		0		NEG		NEG		NEG		NEG	PASS
EGFRv5 19	WT	Unknown	23.298	POS	29.563	POS		NEG		NEG	FAIL
EGFRv5 20	WT	Unknown	24.694	POS		NEG		NEG		NEG	FAIL
EGFRv5 21	WT	Unknown	27.948	POS		NEG	17.898	POS	20.483	POS	FAIL
EGFRv5 22	WT	Unknown	16.599	POS	22.002	POS	18.245	POS	18.875	POS	PASS
EGFRv5 23	Del	Unknown	16.897	POS	22.018	POS		NEG		NEG	FAIL
EGFRv5 24	WT	Unknown	25.133	POS		NEG		NEG		NEG	FAIL
NTC		0		NEG		NEG	39.173	POS		NEG	FAIL
EGFRv5 25	WT	Unknown	17.268	POS	23.712	POS	18.924	POS	22.147	POS	PASS
EGFRv5 26	WT	Unknown	22.769	POS	34.479	POS	28.061	POS		NEG	PASS
EGFRv5 27	WT	Unknown	20.868	POS	28.288	POS	22.299	POS	26.968	POS	PASS
EGFRv5 28	WT	Unknown	22.383	POS	33.795	POS	28.589	POS	33.378	POS	PASS
EGFRv5 29	WT	Unknown	23.426	POS		NEG	30.235	POS	38.301	POS	PASS
EGFRv5 30	WT	Unknown	27.161	POS		NEG		NEG		NEG	FAIL
NTC		0		NEG		NEG		NEG		NEG	PASS
										PCR Pass:	25
										PCR Fail:	9
										Failure rate:	**26.5%**

Table 2.3.5.4 shows the results for the endogenous control assay for all samples tested using the GF EGFR assay version 5. In total 9/ 25 samples failed the endogenous control in at least one master mix, resulting in a failure rate of 26.5%. Of the NTCs, only 1/4 failed which showed an improvement on the results from version 4 of the assay.

By the end of the investigation, 142 FFPE tissue samples had been tested using the five versions of the GF EGFR assays. A summary of the testing is shown in table 2.3.5.5. Overall

49.3% of results matched the original Therascreen results. The assays generally worked better with wild-type samples, which were concordant with the original results in 67.3% of cases, whereas only 38.9% of mutant samples matched the original results. The relatively low sensitivity and specificity of the GF EGFR assays is reflected in the relatively low level of concordance seen in this phase of the project.

Table 2.3.5.5: Summary table showing the concordance % and discrepancy % of all the samples tested with the GeneFirst EGFR PCR assay versions 1- 5. WT = Wild-type.

		GeneFirst EGFR v1 - 5			**Sample Concordance %**	
		Match	Discrepant	Totals	Mutant Samples	WT samples
Therascreen	Mutant	35	55	90	38.9	
	WT	35	17	52		67.3
	Totals	70	72	142		
Total	**Concordance %**	49.3				
	Discrepancy %		50.7			

GENEFIRST *KRAS* ASSAY

2.3.6 GF KRAS Assay Version 1

Unlike version 1 of the EGFR assay, version 1 of the KRAS assay started by utilising multiple master mixes for mutant detection. The assay was designed to detect a total of eight *KRAS*

mutants using two master mixes. The mutations could not be distinguished, so a positive result on a particular master mix indicated that a mutation was present, but did not specify the exact type of mutation. All samples used to test the GeneFirst KRAS Assay were DNA samples extracted from tissue. The mutation detection results of the individual samples tested can be found in table A2.2.1 (appendix).

Table 2.3.6.1 shows a summary of the assays performance. Version 1 of the KRAS assay achieved 88.9% concordance with the original KRAS castPCR results, and 11.1% discrepancy. This performance is significantly better than the performance of version 1 of the EGFR assay. All wild-type samples were correctly identified, although only three such samples were tested with KRAS version 1.

Table 2.3.6.1: Summary table showing the concordance % and discrepancy % of the GeneFirst KRAS PCR assay version 1. WT = Wild-type.

		GeneFirst KRAS v1			**Sample Concordance %**	
		Match	Discrepant	Totals	Mutant Samples	WT samples
CAST	Mutant	13	2	15	86.7	
	WT	6	0	6		100.0
	Totals	19	2	21		
Total	**Concordance %**	90.5				
	Discrepancy %		9.5			

Table 2.3.6.3 Endogenous control results from samples tested using the KRAS assay version 1. N.B. Not all samples analysed using both master mixes (see "PCR MM" column for the master mixes used). MM = Master Mix, NTC = No template control, WT = Wild-type.

Study ID	Mutation type	PCR MM	DNA (ng, total)	ROX Ct (Mean)	Control Pos/ Neg	Assay Pass/ Fail	Comments
KRASv1 01	G12A	A	30.0	27.480	POS	PASS	
KRASv1 02	G12A	A	30.0	38.177	POS	PASS	
KRASv1 03	WT	A	30.0	31.139	POS	PASS	
KRASv1 04	G12A	A	30.0	33.404	POS	PASS	
NTC		A	0.0	20.110	POS	FAIL	NTC ROX positive
KRASv1 08	G12A	A	15.0	13.548	POS	PASS	
KRASv1 09	G12D	A	15.0	23.656	POS	PASS	
KRASv1 10	G12A	A	15.0	21.449	POS	PASS	
KRASv1 11	WT	A	15.0	39.682	POS	PASS	
NTC		A	0.0	20.469	POS	FAIL	NTC ROX positive
KRASv1 12	G12A	A	7.5	25.088	POS	PASS	
KRASv1 13	G12D	A	7.5	29.343	POS	PASS	
KRASv1 14	G12A	A	7.5	29.145	POS	PASS	
KRASv1 15	WT	A	7.5	42.462	POS	PASS	
NTC		A	0.0	20.469	POS	FAIL	NTC ROX positive
KRASv1 05	WT	B	30.0		NEG	FAIL	Endogenous control fail
KRASv1 06	G12C	B	30.0	39.276	POS	PASS	
KRASv1 07	G12V	B	30.0	19.388	POS	PASS	
NTC		B	0.0	22.871	POS	FAIL	NTC ROX positive
KRASv1 16	G12C	B	15.0	31.276	POS	PASS	
KRASv1 17	G12V	B	15.0	26.656	POS	PASS	
KRASv1 18	WT	B	15.0	25.421	POS	PASS	
NTC		B	0.0	11.160	POS	FAIL	NTC ROX positive
KRASv1 19	G12C	B	7.5	33.603	POS	PASS	
KRASv1 20	G12V	B	7.5	28.944	POS	PASS	
KRASv1 21	WT	B	7.5	9.370	POS	PASS	
NTC		B	0.0	11.160	POS	FAIL	NTC ROX positive
					PCR Pass:	20	
					PCR Fail:	7	
					Failure rate:	**25.9%**	

Table 2.3.6.3 shows the ROX endogenous control results for the GF KRAS assay version 1. All 6/6 NTC samples gave a positive signal in the ROX channel, despite there being no template DNA included in any of these reactions. This strongly suggests some form of technical error occurring in one or more of the PCR reagents. Due to these results, testing of this version of the assay was suspended.

2.3.7 GF KRAS Assay Version 2

Version 2 of the GeneFirst KRAS assay did not expand its capabilities to detect new mutations, but the master mixes were redesigned so the mutations were redistributed (see chapter two methods table 2.2.8.2). Table A2.2.2 (appendix) shows all the concordant and discrepant samples tested using version 2. Here more wild-type samples were tested to validate the specificity of the assay.

Table 2.3.7.1 shows a summary of the assays performance. Version 2 of the KRAS assay achieved 63.3% concordance with the original KRAS castPCR results, and 36.7% discrepancy. Version 2 performed significantly worse than version 1. All wild-type samples were correctly identified, displaying 100% concordance with the original KRAS castPCR results. All the discrepant samples came from the mutant samples, either due to the mutations not being detected or additional unexpected mutations would be present.

To test the LoD of the KRAS version 2 assay, positive samples were diluted in the same manner as the EGFR LoD experiments (see section 2.3.2.2). Serial dilutions were prepared, reaching 1% at the low end. The KRAS version 2 assay performed well, detecting both mutants at 0.6 ng of DNA input. Unfortunately there were also some false positive signals detected in the other KRAS PCR master mix, see table 2.3.7.2.

Table 2.3.7.1: Summary table showing the concordance % and discrepancy % of the GeneFirst KRAS PCR assay version 2. WT = Wild-type.

		GeneFirst KRAS v2			Sample Concordance %	
		Match	Discrepant	Totals	Mutant Samples	WT samples
CAST	Mutant	16	12	28	57.1	
	WT	11	0	11		100.0
	Totals	27	12	39		
Total	**Concordance %**	69.2				
	Discrepancy %		30.8			

Table 2.3.7.2: Limit Of Detection (LoD) experiments using mutant samples diluted to varying degrees with WT samples. For each reaction mix: green +/- symbols indicate true results, red +/- symbols indicate false results.

Sample Name	Mutation type	Dilution	DNA (ng, total)	Result (Reaction Mix A)	Result (Reaction Mix D)	Limit of Detection
KRASv2 30	G12V	1	31.4	+	-	**0.6 ng**
		1/10	3.0	+	-	
		1/20	1.5	+	+	
		1/50	0.6	+	-	
KRASv2 43	G13D	1	27.0	-	+	**0.6 ng**
		1/10	3.0	-	+	
		1/20	1.5	-	+	
		1/50	0.6	+	+	

Table 2.3.7.3 Endogenous control results from samples tested using the KRAS assay version 2. N.B. Samples KRASv2 03, 05, 06, 08, 10 were not analysed on both master mixes, therefore corresponding cells without data are indicated with 'N/A' (Not Applicable). ROX A = ROX reporter signal in master mix A, ROX B = ROX reporter signal in master mix B, NTC = No template control, WT = Wild-type.

			ROX A		ROX B			
Study ID	Mutation type	DNA (ng, total)	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Assay Pass/ Fail	Comments
KRASv2 21	G12V	30.0	19.870	POS	21.395	POS	PASS	
KRASv2 07	G12C	30.0	20.219	POS	20.759	POS	PASS	
KRASv2 11	G12V	30.0	20.664	POS	20.785	POS	PASS	
KRASv2 02	WT	30.0	20.468	POS	21.307	POS	PASS	
KRASv2 04	G12A	30.0	20.051	POS	20.766	POS	PASS	
KRASv2 09	G12A	30.0	23.027	POS	23.364	POS	PASS	
KRASv2 23	G12D	30.0	22.397	POS	22.518	POS	PASS	
KRASv2 24	WT	30.0	22.274	POS	22.722	POS	PASS	
NTC		0.0	42.475	(POS)		NEG	PASS	Background only
KRASv2 15	WT	30.0	19.570	POS	20.002	POS	PASS	
KRASv2 16	WT	30.0	22.902	POS	23.126	POS	PASS	
KRASv2 17	WT	30.0	25.109	POS	25.616	POS	PASS	
KRASv2 18	G12D	30.0	21.159	POS	21.649	POS	PASS	
KRASv2 19	G13D	30.0	20.057	POS	20.478	POS	PASS	
KRASv2 20	WT	30.0	21.121	POS	21.487	POS	PASS	
KRASv2 22	WT	30.0	22.000	POS	22.320	POS	PASS	
KRASv2 25	WT	30.0	20.091	POS	20.770	POS	PASS	
KRASv2 26	WT	30.0	20.094	POS	20.458	POS	PASS	
NTC		0.0		NEG		NEG	PASS	
KRASv2 27	G12V	32.6	22.915	POS	23.288	POS	PASS	
KRASv2 28	G13D	27.0	19.814	POS	19.874	POS	PASS	
KRASv2 29	G12D	21.0	23.844	POS	26.716	POS	PASS	
KRASv2 30	G12V	31.4	19.237	POS	19.761	POS	PASS	
NTC		0.0		NEG		NEG	PASS	
KRASv2 31	G13D	25.0	20.602	POS	21.184	POS	PASS	
KRASv2 32	G12V	29.0	19.315	POS	19.675	POS	PASS	
KRASv2 33	G12D	33.2	18.982	POS	19.587	POS	PASS	
KRASv2 34	EGFR WT	37.4	22.163	POS	22.992	POS	PASS	
KRASv2 35	EGFR WT	25.9	21.970	POS	23.954	POS	PASS	
KRASv2 36	EGFR WT	23.0	23.384	POS	22.559	POS	PASS	
KRASv2 37	EGFR WT	31.4	22.010	POS	22.468	POS	PASS	
NTC				NEG		NEG	PASS	
KRASv2 38	WT	30.0	22.286	POS	21.822	POS	PASS	
NTC				NEG	49.202	NEG	PASS	
KRASv2 39	G12V	31.4	18.809	POS	19.139	POS	PASS	
KRASv2 40	G12V	3.0	22.299	POS	21.545	POS	PASS	
KRASv2 41	G12V	1.5	22.521	POS	21.796	POS	PASS	
KRASv2 42	G12V	0.6	22.859	POS	21.996	POS	PASS	
NTC				NEG		NEG	PASS	
KRASv2 43	G13D	27.0	19.115	POS	18.774	POS	PASS	
KRASv2 44	G13D	3.0	20.086	POS	21.341	POS	PASS	
KRASv2 45	G13D	1.5	21.555	POS	21.447	POS	PASS	
KRASv2 46	G13D	0.6	21.879	POS	21.890	POS	PASS	
KRASv2 01	WT	30.0	35.910	POS	23.447	POS	PASS	
KRASv2 03	G12A	30.0		N/A	22.047	POS	PASS	
KRASv2 05	G12D	15.0		N/A	22.088	POS	PASS	
KRASv2 06	G12C	30.0	22.431	POS		N/A	PASS	
KRASv2 08	G12A	30.0	24.163	POS		N/A	PASS	
KRASv2 10	G12V	30.0	21.439	POS		N/A	PASS	
						PCR Pass:	49	
						PCR Fail:	0	
						Failure rate:	**0%**	

Table 2.3.7.3 shows the results of the ROX endogenous control assay for the samples and controls tested using version 2 of the GF KRAS PCR. In five case (cells highlighted with a black border) the samples could only be loaded in one of the two reactions, therefore these reactions were not expected to generate a ROX signal. In all 49 samples and controls tested, all gave the correct result in the endogenous control assay. This is a significant improvement on the performance from version 1.

2.3.8 GF KRAS Assay Version 3

GeneFirst KRAS PCR version 3 improved on the performance of version 2, here achieving 78.9% concordance with the original results, with a discrepancy level of 21.1% (see tables 2.3.8.1 and 2.3.8.2). Again 100% of the wild-type samples matched with the original results. All the discrepant samples came from false positives or false negatives within the mutant positive samples.

Similarly to the GeneFirst EGFR assay, problems with the endogenous control appeared in version 3 of the KRAS assay. On several occasions the No Template Control (NTC) negative control samples started giving a positive signal for the endogenous control (ROX channel). Like the equivalent problem in the EGFR assay, this was initially considered to be accidental user contamination; however successive repeats and reactions prepared with only master mix (i.e. reagents only, no DNA or nuclease free water) continued to produce this endogenous control positive signal. The data from these experiments is shown in the Chapter two appendix tables A2.2.6 and figures A2.2 and A2.3. The source of this signal was later discovered to also be a contamination in the manufacturing process at GeneFirst, so it was not an incident that could be avoided in the UHCW Pathology laboratory.

Table 2.3.8.2: Summary table showing the concordance % and discrepancy % of the GeneFirst KRAS PCR assay version 3. WT = Wild-type.

		GeneFirst KRAS v3			Sample Concordance %	
		Match	Discrepant	Totals	Mutant Samples	WT samples
CAST	Mutant	8	4	12	66.7	
	WT	7	0	7		100.0
	Totals	15	4	19		
Total	**Concordance %**	78.9				
	Discrepancy %		21.1			

Table 2.3.8.3 shows the ROX endogenous control results from the samples and controls tested with the GF KRAS assay version 3. All DNA samples gave a positive ROX signal as expected, however both NTC samples also gave strong (Ct ~15) positive ROX signals. Overall failure rate was 2/19 (9.5%). These results suggest contamination of the PCR reagents.

Table 2.3.8.3 Endogenous control results from samples tested using the KRAS assay version 3. Samples not numbered sequentially due to some samples being excluded for technical reasons. ROX A = ROX reporter signal in master mix A, ROX B = ROX reporter signal in master mix B, NTC = No template control, WT = Wild-type.

Study ID	Mutation type	DNA (ng, total)	ROX A		ROX B		Assay Pass/ Fail
			Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	
KRASv3 01	G12A	6.84	10.135	POS	12.806	POS	PASS
KRASv3 02	G12D	2.14	11.603	POS	13.953	POS	PASS
KRASv3 03	G12C	4.32	10.555	POS	13.087	POS	PASS
KRASv3 04	G12A	4.52	12.065	POS	14.707	POS	PASS
NTC			15.311	POS	17.936	POS	FAIL
KRASv3 06	G12D	2.36	12.458	POS	14.942	POS	PASS
KRASv3 07	G13D	20.80	9.522	POS	12.213	POS	PASS
KRASv3 08	G12D	18.70	10.265	POS	15.089	POS	PASS
KRASv3 09	G12V	21.00	7.844	POS	11.144	POS	PASS
KRASv3 15	G12D	15.00	9.211	POS	12.138	POS	PASS
KRASv3 16	G12V	15.00	12.021	POS	14.485	POS	PASS
KRASv3 17	WT	15.00	11.133	POS	13.686	POS	PASS
KRASv3 18	G13D	15.00	11.002	POS	13.741	POS	PASS
KRASv3 19	WT	15.00	13.768	POS	15.608	POS	PASS
KRASv3 20	WT	15.00	11.811	POS	14.386	POS	PASS
KRASv3 21	WT	15.00	11.942	POS	14.597	POS	PASS
KRASv3 22	WT	15.00	13.513	POS	15.749	POS	PASS
KRASv3 23	WT	15.00	10.268	POS	12.935	POS	PASS
KRASv3 24	G12V	15.00	9.330	POS	12.365	POS	PASS
KRASv3 25	WT	15.00	11.085	POS	13.898	POS	PASS
NTC			15.353	POS	17.697	POS	FAIL
						PCR pass:	19
						PCR fail:	2
						Failure rate:	**9.5%**

2.3.9 GF KRAS Assay Version 4

Only a single run was performed using the version 4 KRAS kit. The samples used all gave very different results to those obtained using castPCR, data shown in table 2.3.9.1. Concordance with castPCR results was low, with only 4 of 15 samples (26.7%) matching. This reduction in assay performance represents a 52.28% fall in concordance from version 3 to version 4 of the KRAS assay. Table 2.3.9.2 shows the overall performance of the assay. In this version of the KRAS assay the wild-type samples performed particularly poorly, with only 1 sample of 9

(11%) matching previous results using the castPCR assay. Only 6 mutant samples were available at this phase of the project. 3 of the 6 (50%) mutant samples matched previous results.

Table 2.3.9.2: Summary table showing the concordance % and discrepancy % of the GeneFirst KRAS PCR assay version 4. WT = Wild-type.

		GeneFirst KRAS v4			**Sample Concordance %**	
		Match	Discrepant	Totals	Mutant Samples	WT samples
CAST	Mutant	3	3	6	50.0	
	WT	1	8	9		11.1
	Totals	4	11	15		
Total	**Concordance %**	26.7				
	Discrepancy %		73.3			

Table 2.3.9.3 Endogenous control results from samples tested using the KRAS assay version 4. ROX A = ROX reporter signal in master mix A, ROX B = ROX reporter signal in master mix B, NTC = No template control, WT = Wild-type.

Study ID	Mutation type	DNA (ng, total)	ROX A		ROX B		Assay Pass/ Fail
			Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	
KRASv4 01	G12D	24.4	20.956	POS	28.177	POS	PASS
KRASv4 02	WT	5.4	23.838	POS		NEG	FAIL
KRASv4 03	WT	10.1	22.562	POS	30.567	POS	PASS
KRASv4 04	WT	9.3	21.173	POS	28.425	POS	PASS
KRASv4 05	WT	5.6	23.301	POS	33.565	POS	PASS
KRASv4 06	WT	0.7	25.277	POS		NEG	FAIL
KRASv4 07	WT	11.4	20.985	POS	29.964	POS	PASS
KRASv4 08	G12A	8.7	20.577	POS	28.669	POS	PASS
KRASv4 09	G12C	7.2	21.123	POS	28.908	POS	PASS
KRASv4 10	G12V	16.0	20.556	POS	28.847	POS	PASS
KRASv4 11	WT	0.4	26.185	POS		NEG	FAIL
KRASv4 12	WT	1.0	29.678	POS		NEG	FAIL
KRASv4 13	WT	3.3	21.809	POS	29.809	POS	PASS
KRASv4 14	G12V	4.2	22.114	POS	30.511	POS	PASS
KRASv4 15	G12D	4.2	23.548	POS	30.737	POS	PASS
NTC				NEG		NEG	PASS
NTC				NEG	39.728	POS	FAIL
NTC				NEG	39.401	POS	FAIL
						PCR Pass:	12
						PCR Fail:	6
						Failure rate:	**33.3%**

Table 2.3.9.3 shows the performance of the ROX endogenous control assay from GF KRAS PCR version 4. For master mix A all samples and controls performed as expected. In master mix B however four samples gave a negative result in the ROX channel. This could potentially be due to low DNA input, as the quantity of DNA loaded for these reactions ranged from only 0.4- 5.4 ng, below the guideline concentration of 30 ng per reaction. Conversely two of three NTC samples gave a positive ROX signal in master mix B, however the Ct values were high (39.728 and 39.401) suggesting a weak signal.

2.3.10 GF KRAS Assay Version 5

18 samples were tested using the final version of the GeneFirst KRAS assay. Table 2.3.10.1 shows a summary of the results for all samples tested. A major improvement in concordance observed compared to version 4: 66.7% in version 5 compared to 26.7% in version 4. Table 2.3.10.2 shows the overall performance of the assay. The assay performed equally well with wild-type and mutant samples, where 66.7% of both sample types matched the results from the castPCR assay.

Table 2.3.10.1: Summary table showing the concordance % and discrepancy % of the GeneFirst KRAS PCR assay version 5. WT = Wild-type.

		GeneFirst KRAS v5			**Sample Concordance %**	
		Match	Discrepant	Totals	Mutant Samples	WT samples
CAST	Mutant	8	4	12	66.7	
	WT	4	2	6		66.7
	Totals	12	6	18		
Total	**Concordance %**	66.7				
	Discrepancy %		33.3			

To determine the exact sensitivity and specificity of the KRAS version 5 assay, all the PCR data was analysed, including the results from the individual master mixes. The KRAS version 5 assay utilised two master mixes, each with two reporters (VIC and FAM) so therefore it was possible to generate up to four signals per sample. The results are shown in table 2.3.10.2.

In all cases, the discrepant results were false positives i.e. one or more master mixes gave a signal that corresponded to a mutation that had not been previously been detected by the castPCR assay. This phenomena affected both master mixes equally, with six false positive results in master mix A, and five false positive results in master mix B.

Table 2.3.10.2: Complete analysis of PCR results from all GF KRASv5 master mixes. Designations 'True Positive', 'True Negative', 'False Positive' and 'False Negative' are assigned based on the original castPCR results. WT = Wild-type.

Study ID	Mutation type	GF KRASv5 PCR Master Mix		Result			
		A	B	True Pos	True Neg	False Pos	False Neg
KRASv5 01	WT	TRUE NEG	TRUE NEG	0	2	0	0
KRASv5 02	WT	TRUE NEG	TRUE NEG	0	2	0	0
KRASv5 03	G12V	TRUE POS	TRUE NEG	1	1	0	0
KRASv5 04	G12C	TRUE POS	TRUE NEG	1	1	0	0
KRASv5 05	G12C	TRUE POS	TRUE NEG	1	1	0	0
KRASv5 06	G12D	FALSE POS	FALSE POS	0	0	2	0
KRASv5 07	G12D	TRUE POS	TRUE POS	2	0	0	0
KRASv5 08	G12A	FALSE POS	TRUE POS	1	0	1	0
KRASv5 09	WT	FALSE POS	FALSE POS	0	0	2	0
KRASv5 10	WT	TRUE NEG	TRUE NEG	0	2	0	0
KRASv5 11	WT	TRUE NEG	TRUE NEG	0	2	0	0
KRASv5 12	G12V	FALSE POS	FALSE POS	0	0	2	0
KRASv5 13	G12C	TRUE POS	TRUE NEG	1	1	0	0
KRASv5 14	G12C	TRUE POS	TRUE NEG	1	1	0	0
KRASv5 15	G12D	TRUE POS	TRUE POS	2	0	0	0
KRASv5 16	G12D	TRUE POS	TRUE POS	2	0	0	0
KRASv5 17	G12A	FALSE POS	FALSE POS	0	0	2	0
KRASv5 18	WT	FALSE POS	FALSE POS	0	0	2	0
Totals:	**TRUE**	**12**	**13**	**12**	**13**	**11**	**0**
	FALSE	**6**	**5**				

The analysis from table 2.3.10.2 was used calculate the sensitivity and specificity of the GF KRAS version 5 assay. The results are shown in table 3.3.10.3. The assay achieved 100% sensitivity, but only 54.2% specificity.

Table 2.3.10.3 Endogenous control results from samples tested using the KRAS assay version 5. ROX A = ROX reporter signal in master mix A, ROX B = ROX reporter signal in master mix B, NTC = No template control, WT = Wild-type.

Study ID	Mutation type	DNA (ng, total)	ROX A		ROX B		Assay Pass/ Fail
			Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	
KRASv5 01	WT	30.0		NEG	29.629	POS	FAIL
KRASv5 02	WT	30.0	26.220	POS	26.285	POS	PASS
KRASv5 03	G12V	30.0	20.440	POS	21.520	POS	PASS
KRASv5 04	G12C	30.0	22.532	POS	23.735	POS	PASS
KRASv5 05	G12C	30.0	29.616	POS		NEG	FAIL
KRASv5 06	G12D	30.0	22.093	POS	21.418	POS	PASS
KRASv5 07	G12D	30.0	24.379	POS	24.644	POS	PASS
KRASv5 08	G12A	30.0	21.755	POS	20.911	POS	PASS
KRASv5 09	WT	30.0	28.380	POS	25.461	POS	PASS
NTC		0.0		NEG		NEG	PASS
KRASv5 10	WT	15.0	29.965	POS	26.412	POS	PASS
KRASv5 11	WT	15.0	25.939	POS	25.528	POS	PASS
KRASv5 12	G12V	15.0	21.510	POS	22.580	POS	PASS
KRASv5 13	G12C	15.0	21.921	POS	22.611	POS	PASS
KRASv5 14	G12C	15.0	27.893	POS	27.188	POS	PASS
KRASv5 15	G12D	15.0	22.266	POS	22.074	POS	PASS
KRASv5 16	G12D	15.0	25.269	POS	24.932	POS	PASS
KRASv5 17	G12A	15.0	22.067	POS	21.438	POS	PASS
KRASv5 18	WT	15.0	26.925	POS	25.510	POS	PASS
NTC		0.0		NEG		NEG	PASS
						PCR Pass:	18
						PCR Fail:	2
						Failure rate:	**10.0%**

Table 2.3.10.3 shows the performance of the ROX endogenous control in GF KRAS version 5 assay. Two of the 20 samples failed, resulting in a 10% failure rate. Both the NTCs gave no signal in the ROX channel which was the expected result. This is an improvement on version 4 of the assay, where overall 33% of samples failed the endogenous control assay, including 2/3 of the NTCs.

Table 2.3.10.4: Summary of GF KRAS version 5 data.

Gene First KRAS Assay version 5				
Test	**Mutant Present**		**Mutant absent**	
Positive	True Positive	12	False Positive	11
Negative	False Negative	0	True Negative	13

Statistic	**Value**
Sensitivity	**100.0%**
Specificity	**54.2%**
Positive Predictive Value	**52.2%**
Negative Predictive Value	**100.0%**

In total 112 samples were tested using the five different versions of the GF KRAS assay. The data concordance is shown in table 2.3.10.5. Overall 77 of the 112 (68.8%) samples tested matched the original castPCR results. The assays performed slightly better with wild-type samples, as 74.4% of wild-type samples matched the original castPCR results, compared to 65.8% of matching mutant samples.

Table 2.3.10.5: Summary table showing the overall concordance % and discrepancy % of the GeneFirst KRAS PCR assay versions 1-5. WT = Wild-type.

		Gene First KRAS v1-5			Sample Concordance %	
		Match	Discrepant	Totals	Mutant Samples	WT samples
CAST	Mutant	48	25	73	65.8	
	WT	29	10	39		74.4
	Totals	77	35	112		
Total	**Concordance %**	68.8				
	Discrepancy %		31.3			

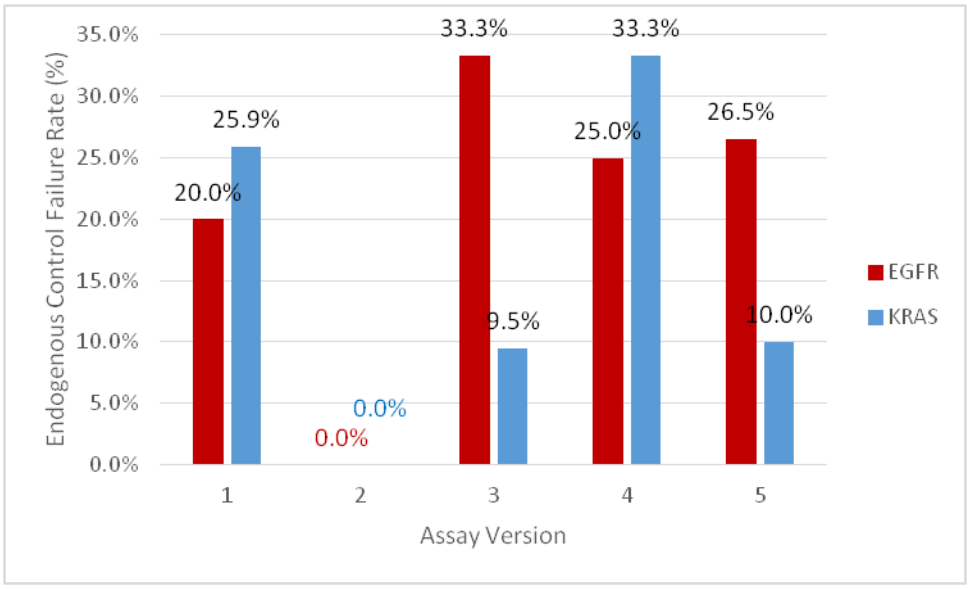

Figure 2.3.10.1 Summary graph of EGFR and KRAS Endogenous Control performance across all tested versions. Data from the EGFR assay is coloured red, data from the KRAS assay is coloured blue.

Figure 2.3.10.1 shows the frequency of endogenous control failure across several versions of both prototype assays. From the graph it can be clearly seen that failure rate varied greatly.

For EGFR the failure rate ranged from 33.3% (version 3) to 0% (version 2). For KRAS the failure rate ranged from 33.3% (version 4) to 0% (version 2). Interestingly version 2 of both assays had a 0% failure rate for endogenous control, but unfortunately this performance could not be replicated for future versions.

2.3.11 Confirmatory Testing using Next Generation Sequencing

Due to the significant number of discrepant samples identified by the validation of the GeneFirst assays using previously tested Pathology samples, Next Generation Sequencing (NGS) was utilised to establish the exact genotype of a selection of the discrepant samples.

The samples that were selected for NGS analysis were selected based on the fact they had been tested and did not match the results determined by either Therascreen or castPCR, and also had a sufficient remaining volume of DNA extract for analysis. The samples were all DNA samples extracted from tissue.

A total of nine samples were sequenced over three runs using the Ion Torrent PGM platform. Amplified DNA libraries were prepared from the DNA extracts using the primer pool from the 22-gene panel (later incorporated into the Oncomine Solid Tumour DNA kit), as described in the Methods (section 2.2.8.3).

The hotspot mutations detected in *EGFR* and *KRAS* are shown in table 2.3.11.1. In total nine *EGFR* or *KRAS* hotspot mutations were detected in seven of the nine samples: three in *EGFR*; six in *KRAS*. Two samples (EGFRv2 46 and EGFRv3 06) were wild-type for *EGFR* and *KRAS*. Of the nine hotspots detected, there were two (*EGFR* p.P741L and *EGFR* p.A871T) that the

GeneFirst EGFR assay was not designed to detect, therefore I would not expect these mutations to give a positive result in the GeneFirst EGFR assay.

Table 2.3.11.1: Hotspot mutations detected in samples from the GeneFirst EGFR and KRAS assays.

Study ID	Frequency	Quality	Gene ID	Allele Name	AA change	Coverage	Allele Cov	Strand Bias
EGFRv2 04	88.6	12667.50	EGFR	COSM6224	p.L858R	1996	1768	0.5073
EGFRv2 43	2.8	12.52	EGFR	COSM17570	p.P741L	177	5	0.5527
EGFRv2 43	41.3	774.99	KRAS	COSM522	p.G12A	213	88	0.5634
EGFRv2 44	2.3	8.55	EGFR	COSM28605	p.A871T	133	3	0.5617
EGFRv2 44	25.8	401.81	KRAS	COSM516	p.G12C	229	59	0.5199
EGFRv2 46	*No EGFR or KRAS Hotspot Mutations Detected*							
EGFRv3 06	*No EGFR or KRAS Hotspot Mutations Detected*							
KRASv2 03	40.3	7238.55	KRAS	COSM522	p.G12A	2000	806	0.5094
KRASv2 05	3.5	9.21	KRAS	COSM521	p.G12D	260	9	0.7188
KRASv2 07	48	4827.25	KRAS	COSM516	p.G12C	1019	489	0.5117
KRASv3 08	50.3	7922.22	KRAS	COSM521	p.G12D	1546	777	0.5096

The NGS results were compared to corresponding data generated by the Therascreen, castPCR and GeneFirst assays. The results are shown in table 2.3.11.2.

Table 2.3.11.2 Mutation detection results from five assays: Therascreen, GeneFirst EGFR PCR, castPCR, GeneFirst KRAS PCR and Ion Torrent NGS. WT = Wild-type.

Study ID	Therascreen Result (EGFR)	GF EGFR Result	CastPCR Result (KRAS)	GF KRAS Result	NGS Result	Comments
EGFRv2 04	L858R	T790M, L858R, Exon 19 Del	N/A	N/A	EGFR p.L858R	GF does not match NGS
EGFRv2 43	WT	T790M	N/A	G12A	KRAS p.G12A	GF matched NGS KRAS result but not EGFR
EGFRv2 44	WT	T790M	N/A	G12C	KRAS p.G12C	GF matched NGS KRAS result but not EGFR
EGFRv2 46	WT	WT	N/A	WT	WT	All tests match
EGFRv3 06	WT	WT	N/A	N/A	WT	All tests match
KRASv2 03	N/A	N/A	G12A	G13D	KRAS p.G12A	GF does not match NGS
KRASv2 05	N/A	N/A	G12D	G13D	KRAS p.G12D	GF does not match NGS
KRASv2 07	N/A	N/A	G12C	WT	KRAS p.G12C	GF does not match NGS
KRASv3 08	N/A	N/A	G12D	G12D and G12A/C/V	KRAS p.G12D	GF does not match NGS

In all cases, the NGS sequencing results agreed with the original Qiagen Therascreen EGFR RGQ PCR and the Life Technologies KRAS castPCR assays. In total 7 false positive mutation results were identified amongst six samples tested on the GeneFirst assays. In one sample the GeneFirst KRAS PCR identified it as WT, but the presence of a G12C mutation was confirmed by NGS. Therefore that sample was a false negative according the GeneFirst KRAS PCR result.

There were two additional *KRAS* mutations detected by the GeneFirst assays that were confirmed by the NGS analysis (see samples EGFRv2 43 and EGFRv2 44). Two wild-type samples were also sequenced to act as negative controls (EGFRv2 46 and EGFRv4 06). The NGS results for these two samples matched the negative results from Therascreen, no *EGFR* (or *KRAS*) mutations were detected.

EGFR samples 43, 44 and 46 were also tested using the GF *KRAS* assay, two of which gave positive signals for *KRAS* mutations. Interestingly NGS analysis confirmed all three of these results. This could suggest that the GF *KRAS* assay actually performed slightly better than the GF EGFR assay. However this observation cannot be confirmed with such a limited number of samples analysed by NGS.

In addition, a wide range of additional mutations and novel variants were identified in the samples which would not have been detectable by the PCR methods. The full range of hotspots identified are given in the chapter two appendix, table A2.3.

Unfortunately due to DNA samples being depleted during the various PCR testing phases, no further samples could be analysed by NGS. Therefore the results from versions 4 and 5 of the GeneFirst EGFR and KRAS assays could not be confirmed using the Ion Torrent platform.

2.3.12 GF EGFR/ KRAS Assays and cfDNA

To observe the performance of the GeneFirst PCR assays with cfDNA, 5 cfDNA samples were analysed using the GF EGFR (version 3) and GF KRAS (version 3) assays. For details of the cfDNA samples and their preparation, see chapter four Methods (sections 4.2.3.1- 3). Due to the highly variable DNA concentrations in the cfDNA samples, they could not all be loaded into the PCR reactions at the same concentration. Table 2.3.12.3 shows the PCR and corresponding NGS results for the 5 cfDNA samples.

Table 2.3.12.1 Endogenous control results from plasma samples tested on EGFR assay version 3. ROX A = ROX reporter signal in master mix A, ROX B = ROX reporter signal in master mix B, ROX C = ROX reporter signal in master mix C, ROX D = ROX reporter signal in master mix D, NTC = No template control, WT = Wild-type.

			ROX A		ROX B		ROX C		ROX D		
Study ID	Mutation type	DNA (ng, total)	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Assay PASS/ FAIL
PL001	WT	7.4	23.746	POS	23.808	POS	21.317	POS	22.635	POS	PASS
PL002	WT	2.2	25.205	POS	25.517	POS	24.012	POS	25.109	POS	PASS
PL005	WT	1.0	28.722	POS	28.463	POS	28.731	POS	31.450	POS	PASS
PL003	WT	1.1	29.372	POS	28.886	POS	26.553	POS	27.793	POS	PASS
PL004	WT	0.9	30.003	POS	29.576	POS	27.554	POS	27.533	POS	PASS
NTC		0.0	31.221	POS	30.389	POS	28.150	POS		NEG	FAIL
										PCR Pass:	5
										PCR Fail:	1
										Failure rate:	33.3%

Table 2.3.12.2 Endogenous control results from plasma samples tested on *KRAS* assay version 3. ROX A = ROX reporter signal in master mix A, ROX B = ROX reporter signal in master mix B, NTC = No template control, WT = Wild-type.

			ROX A		ROX B		
Study ID	Mutation type	DNA (ng, total)	Ct Mean	Control Pos/ Neg	Ct Mean	Control Pos/ Neg	Assay Pass/ Fail
PL001	WT	7.4	11.604	POS	13.660	POS	PASS
PL002	WT	2.2	12.733	POS	15.285	POS	PASS
PL005	WT	1.0	15.944	POS	18.192	POS	PASS
PL003	WT	1.1	15.917	POS	17.990	POS	PASS
PL004	WT	0.9	16.471	POS	18.405	POS	PASS
NTC		0.0	14.987	POS	17.489	POS	FAIL
						PCR Pass:	5
						PCR Fail:	1
						Failure rate:	33.3%

Table 2.3.12.3 cfDNA samples analysed with GF PCR and NGS.

Plasma Sample No.	GeneFirst PCR Results		NGS Results		Concordance	
	EGFR	KRAS	Variants	Hotspots	EGFR	KRAS
PL001	T790M POS	NEG	24	0	Discrepant	Match
PL002	NEG	NEG	37	0	Match	Match
PL003	NEG	NEG	41	0	Match	Match
PL004	NEG	NEG	42	0	Match	Match
PL005	T790M POS	NEG	48	0	Discrepant	Match

Analysing the cfDNA samples by EGFR PCR has detected T790M mutations in two samples. However these mutations were not detected in the same samples using NGS. The PCR results from the KRAS assay agreed with the NGS results, both assays found the cfDNA samples to be negative for *KRAS* mutations.

2.3 DISCUSSION

2.3.1 GeneFirst EGFR PCR assay development using Tissue

Five versions of the GeneFirst EGFR PCR assay were tested over the course of this phase of the project (Sept 2013- December 2014). In the final version, the assay achieved an overall concordance of 49.3% (see table 2.3.5.1) compared to the original Pathology Department results using Qiagen Therascreen EGFR assay. These results are clearly not suitable for use as a clinical diagnostic assay. An earlier version of the assay (version 2) did achieve 1- 2.5% LoD and was able to perform analysis on very low gDNA input (<1 ng in some cases). The decision was made not to perform any further LoD experiments until the issue of frequent false positives was resolved in the assay, hence only version 2 was used for LoD experiments.

The most significant issue was the frequent occurrence of false positives. In some cases these false positives may have been due to DNA cross-linking due to DNA fixation with formalin or probe degradation, however the level of discrepancy (>50%) is too high to be explained just by formalin DNA changes and PCR artefacts. This problem needed to be overcome in further developmental steps in assay design, but the technology has now been abandoned by GeneFirst.

In addition to the false positive error rate encountered with all versions of the assays, there was also an incidence of reagent contamination at the manufacturing stage. The reagents for version 3 of the EGFR assay produced results where all reactions (including non-template negative controls) gave positive results for the control assay. The source of this signal was later discovered to be a contamination in the manufacturing process at GeneFirst.

Under normal circumstances, data generated from runs with failed controls would be excluded from a validation dataset. However in the case of this series of experiments, the general performance of the assay was such that including this sub-optimal data did not have a huge impact on the overall quality of the dataset.

Version 4 of the assay gave a minor improvement in overall assay concordance (46.7%) with the original results, compared to version 3 (39.1%). 100% Wild-type samples matched original results, however there were only 2 wild-type samples tested, so this result is too underpowered to be considered a positive reflection of assay performance.

Unfortunately with the following version 5, assay performance again showed very low concordance with previous results (26.7%). This could have been partially due to a number of very low DNA concentration samples being used, however a majority of these low DNA concentration samples were previously identified as wild-type, and thus I would not expect them to give a positive signal for any mutants. Performance with mutant samples was the worst of all assay versions (10.5% concordance). A major weakness of the assay was shown to be PCR master mix D (designed to detect exon 19 deletions and the L858R mutation) which had the most frequent occurrence of false positives, almost as many as the other three master mixes- 19 false positives in master mix D alone, compared to 21 in master mixes A, B and C combined. Without knowing the exact technical specifications of the design of the primer and probes used in this master mix it is impossible to state exactly why this problem occurred so frequently, however it was most likely due to low specificity of the assay for its intended mutant targets. Alternatively it could have been due to changes in the assay chemistry, or poor sample quality, or a combination of both factors. Analysis of the results was also very cumbersome due to the varying thresholds and multiple reporters used in the four PCR master mixes.

Similar to the variable performance of the mutant target assays, the endogenous control assay included in the GF EGFR PCR regularly failed during the testing phases. The best results for this assay component were seen in version two, were the endogenous control performed correctly for all samples. All other versions however suffered some degree of failure, with the worse performance being in version three where the assays failed in 33.3% of the samples. This unreliable performance of the endogenous control assay further shows that these prototype assays are not suitable for diagnostic use. In a number of cases, the failure of many samples and controls in the endogenous control assay may have been due to contamination of the reagents at the manufacturing stage (explained earlier). This is the most likely explanation of the frequent phenomena of non-template controls giving positive ROX signals. Beyond this however analysis of the endogenous control performance is difficult as no technical details of the endogenous control assay (primer design etc.) were provided by the manufacturer. Analysis of the PCR products using gel electrophoresis (see figure 2.3.3) did not show any evidence of primer dimers or amplicons (specific or non-specific), however this isn't necessarily a definitive answer as the PCR products may have contained fragments at a concentration below that which can be visualised by the gel.

2.3.2 GeneFirst KRAS PCR assay development using Tissue

Five versions of the GeneFirst KRAS PCR assay were tested over the course of the project. Generally this assay performed better than the EGFR assay, achieving a concordance of 68.8% across all versions of the assay (see table 2.3.10.5), compared to 49.3% for the EGFR assay overall (see table 2.3.5.5). Version 2 of the assay achieved a good LoD at 1% in the LoD experiments, and was able to perform valid analysis using low inputs of gDNA (7.5 ng). The

advantage for the *KRAS* testing was that the original Pathology Department samples were extracted using punched sections from tissue blocks, rather than the scraped sides used for the EGFR assay, so that much higher gDNA yields and sample volumes were available.

Similarly to the EGFR assay, the most significant source of error was the occurrence of false positives amongst mutation positive samples. Most often the false positive result was caused by anomalous signals in the FAM channel in one of the master mixes, which in combination with a similar signal in the VIC channel gave a positive result for a specific mutation (Gly13Asp). This false positive occurred frequently in version 2 of the assay (see NGS data table 2.3.11.2 and appendix table A2.2.2) and is most likely a feature of sub-optimal PCR conditions, rather than DNA cross linking or probe degradation, as described in the previous paragraph. Despite these issues, the KRAS assay performed better than the EGFR assay.

In addition to the high false positive rate, the other major problem was the manufacturing contamination incident detected in version 3 of both kits. Although this does not represent an innate problem with the assay, more an unfortunate incident during manufacture, it did however invalidate the testing done on version 3 and represent a considerable amount of time lost.

Versions 4 of the KRAS assays saw considerable reductions in performance. Version 4 achieved an overall concordance of only 26.7%, compared to 79% in the previous version (version 3, although these reagents were contaminated at source so the results are not reliable).

Version 5 achieved an improvement over version 4 (66.7% compared to 26.7%) however there was still very high incidence of false positive results, although with the limited number

of samples available it was not possible to identify if this was an issue associated with either of the master mixes. One major improvement in version 5 saw the assay generate no false negatives, resulting in a sensitivity of 100% for this version of the assay. Specificity unfortunately was low at 54.2% due the frequent occurrences of false positives. Like the EGFR assay, it is impossible to know the exact cause of these issues without technical information about the assay design. This could have been due to changes in the assay chemistry or poor sample quality.

Another issue that arose with version 5 was the introduction of a very complex and confusing analysis algorithm in the kit SOP (see Methods figure 2.8.2). Not only was the analysis flow diagram difficult to interpret, but it also gave imprecise thresholds for determining ΔCt cut offs for separating positive and negative results. In a number of cases, symbols such as << and >> (without associated value ranges) were given as criteria for determining the PCR results instead of discreet thresholds, so therefore some interpretation and subjectivity was employed by the operator. This process was variable and cumbersome, and far too prone to human error to be suitable for a diagnostic assay.

Endogenous control performance across all five versions of the GF KRAS assay was highly variable. The best results for this assay component were seen in version two, were the endogenous control performed correctly for all samples. All other versions however suffered some degree of failure, with the worse performance being in version four where the assays failed in 33.3% of the samples. This unreliable performance of the endogenous control assay further shows that these prototype assays are not suitable for diagnostic use. Similar to the mutant target assays, no technical information regarding the endogenous control assay was supplied by GeneFirst, therefore it is difficult to explain exactly why this assay gave such variable results.

2.3.3 NGS application

Next Generation Sequencing was used in a limited capacity to perform confirmatory testing on discrepant samples from the GeneFirst EGFR and KRAS PCR assays. This technology provides a "gold standard" in terms of accuracy, sensitivity and depth of coverage for genetic analysis. However due to significantly higher cost of equipment and reagents, and the more time consuming and labour intensive workflow, it was not at the time a suitable method for large scale testing in patient samples within pathology.

NGS confirmatory sequencing showed that only one of the five *EGFR* mutations identified by the EGFR PCR assay were correct, whilst the other 4 were not detected by Qiagen Therascreen (tested by Pathology Department) or by NGS. This strongly suggests that the additional mutations identified by the EGFR assay were false positives.

Seven of the nine NGS analysed samples were analysed using the GeneFirst KRAS assay, identifying six mutations in total. The NGS data agreed with three of the six mutations found. Interestingly the NGS analysis identified two *KRAS* mutations in two samples that had been previously only analysed for *EGFR* mutations. This finding makes a strong case for wider genetic analysis of tumour tissue samples, the mutations identified would have clinical implications and thus this would a useful tool for patient management.

In addition to confirming the presence or absence of mutations identified by PCR, the NGS technology also identified a considerable number of other mutations and novel variants present in the samples (see appendix table A2.3). The significance of these additional mutants and variants need to be investigated using various online genetic reference libraries.

The detection of such a range of genetic features further demonstrated the power of the NGS technologies over the PCR format assays.

2.3.4 Conclusion

The GeneFirst EGFR and KRAS PCR assays demonstrated good levels of sensitivity at certain points during the investigation, although this varied greatly as the assay versions changed. The best examples are the LoD experiments, and the final version of the KRAS assay demonstrated 100% sensitivity. However the specificity of this technical approach needs to be significantly improved before it can be considered for diagnostic use. Manufacturing issues were apparent, although these may have been isolated incidences. Another barrier to the diagnostic use of these assays (assuming the specificity issue were corrected) is the complexity of the analysis process. The equations for calculating ΔCt were straightforward and could most likely be automated, however the extremely confusing and complicated flow diagram for determining the final results (introduced in KRAS version 5) made the analysis very cumbersome and extremely prone to misinterpretation. If this format of assay were to ever be considered for diagnostic use then a more user friendly analysis algorithm would have to be designed. With all the above observations considered, both GeneFirst assays were shown to not be fit for purpose.

The Ion Torrent NGS technology showed itself to be powerful, and proved to be an extremely useful tool for genetic analysis on a selected number of samples during this phase of the project. Unfortunately it was not possible to analyse all discrepant samples identified by the GeneFirst PCR assays by NGS. There were two reasons for this: firstly, the cost of NGS

reagents meant only small number of chips could be allocated to this section of the project; and secondly, small volume of sample DNA available often meant repeats were not possible.

The major limitation of the NGS technology was cost and as a result it saw only limited use here. It will be employed more in following chapters. PCR technologies are very efficient in terms of cost and as a result it is preferable to develop PCR based diagnostic tools rather than NGS based. Assays like the GeneFirst prototype EGFR and KRAS assays could be appropriate for diagnostic use but only after very extensive development to increase the specificity and to streamline the analysis process.

A2 CHAPTER TWO APPENDIX

A2.1 GeneFirst EGFR raw data versions 1-5

Table A2.1.1: Anonymised Pathology samples tested using the GeneFirst EGFR PCR assay version 1. Positive concordance value (%) shown in bottom right cell. WT = Wild-type.

Study ID	Mutations	DNA Loading (ng)	Concordance with Therascreen
EGFRv1 17	Exon 19 Del	52.56	Match
EGFRv1 14	WT	34.60	Discrepant
EGFRv1 16	T790M	34.02	Match
EGFRv1 08	T790M	33.40	Match
EGFRv1 18	T790M	26.18	Discrepant
EGFRv1 05	L858R	25.60	Discrepant
EGFRv1 12	WT	25.20	Match
EGFRv1 09	T790M	18.40	Match
EGFRv1 03	Exon 19 Del	16.80	Match
EGFRv1 07	L858R	16.80	Match
EGFRv1 15	T790M & L858R	16.80	Match
EGFRv1 02	Exon 19 Del	16.00	Match
EGFRv1 06	L858R	15.80	Discrepant
EGFRv1 21	S768I	13.48	Match
EGFRv1 10	WT	12.00	Discrepant
EGFRv1 04	Exon 19 Del	11.20	Match
EGFRv1 11	WT	9.00	Match
EGFRv1 01	Exon 19 Del	7.80	Match
EGFRv1 13	WT	6.00	Match
EGFRv1 20	Exon 19 Del	2.50	Discrepant
EGFRv1 19	L861Q	1.64	Discrepant
Mean PCR DNA input (ng)	Match:	**19.82**	**Concordance**
	Discrepant:	**16.90**	**66.67%**

Table A2.1.2: Anonymised Pathology samples tested using the GeneFirst EGFR PCR assay version 2. Positive concordance percentage shown in the bottom right cell. WT = Wild-type.

Study ID	Mutations	DNA Loading (ng)	Concordance with Therascreen	Study ID	Mutations	DNA Loading (ng)	Concordance with Therascreen
EGFRv2 13	T790M	80.00	Discrepant	EGFRv2 03	Exon 19 Del	8.40	Match
EGFRv2 32	WT	73.92	Match	EGFRv2 05	L858R	8.40	Discrepant
EGFRv2 17	Exon 19 Del	56.96	Discrepant	EGFRv2 02	Exon 19 Del	8.00	Match
EGFRv2 11	T790M	39.27	Discrepant	EGFRv2 08	WT	4.50	Match
EGFRv2 29	WT	39.20	Match	EGFRv2 34	WT	4.14	Match
EGFRv2 43	WT	37.40	Discrepant	EGFRv2 53	T790M	3.00	Match
EGFRv2 19	WT	34.60	Discrepant	EGFRv2 28	WT	2.90	Match
EGFRv2 46	WT	31.40	Match	EGFRv2 30	WT	2.88	Match
EGFRv2 47	WT	30.00	Match	EGFRv2 48	L858R	2.80	Discrepant
EGFRv2 52	T790M	30.00	Discrepant	EGFRv2 40	WT	2.34	Match
EGFRv2 44	WT	25.90	Discrepant	EGFRv2 38	Exon 19 Del	1.64	Discrepant
EGFRv2 15	L858R	25.60	Discrepant	EGFRv2 49	L858R	1.50	Discrepant
EGFRv2 45	WT	23.00	Discrepant	EGFRv2 54	T790M	1.50	Match
EGFRv2 33	WT	22.56	Match	EGFRv2 42	WT	1.39	Discrepant
EGFRv2 14	T790M	18.40	Discrepant	EGFRv2 39	E18 G719X	1.06	Match
EGFRv2 27	WT	18.00	Discrepant	EGFRv2 50	L858R	1.00	Discrepant
EGFRv2 41	WT	17.68	Match	EGFRv2 12	L861Q	0.82	Match
EGFRv2 16	L858R	16.80	Match	EGFRv2 55	T790M	0.60	Match
EGFRv2 01	Exon 19 Del	16.00	Match	EGFRv2 31	WT	0.52	Match
EGFRv2 20	L858R	15.80	Match	EGFRv2 51	L858R	0.50	Match
EGFRv2 25	WT	15.32	Match	EGFRv2 22	WT	0.48	Match
EGFRv2 18	S768I	13.80	Match	EGFRv2 37	Exon 19 Del	0.43	Discrepant
EGFRv2 10	T790M	13.09	Discrepant	EGFRv2 23	WT	0.40	Match
EGFRv2 04	L858R	12.80	Discrepant	EGFRv2 24	WT	0.40	Match
EGFRv2 09	WT	12.60	Match	EGFRv2 36	WT	0.36	Match
EGFRv2 21	WT	12.00	Match	EGFRv2 26	WT	0.32	Discrepant
EGFRv2 35	WT	9.86	Match	**Mean PCR DNA Input (ng)**	Match:	**15.07**	**Concordance**
EGFRv2 06	T790M	9.20	Discrepant		Discrepant:	**14.71**	**60.38%**
EGFRv2 07	T790M	9.20	Discrepant				

Table A2.1.3: Anonymised Pathology samples tested using the GeneFirst EGFR PCR assay version 3. Positive concordance value (%) shown in bottom right cell. WT = Wild-type.

Study ID	Mutations	DNA Loading (ng)	Concordance with Therascreen
EGFRv3 11	WT	207.20	Match
EGFRv3 06	T790M	160.00	Discrepant
EGFRv3 08	WT	149.60	Discrepant
EGFRv3 05	Exon 19 Del	113.92	Discrepant
EGFRv3 09	WT	74.80	Discrepant
EGFRv3 14	WT	69.20	Match
EGFRv3 23	L861Q	60.80	Discrepant
EGFRv3 27	L861Q	60.80	Match
EGFRv3 01	T790M	60.00	Match
EGFRv3 12	WT	51.80	Match
EGFRv3 02	T790M	36.80	Match
EGFRv3 07	T790M	26.18	Discrepant
EGFRv3 03	T790M	18.40	Discrepant
EGFRv3 26	Exon 19 Del	18.24	Discrepant
EGFRv3 15	L858R	16.80	Discrepant
EGFRv3 24	L858R	16.24	Discrepant
EGFRv3 10	WT	10.00	Discrepant
EGFRv3 04	L858R	5.60	Discrepant
EGFRv3 22	Exon 19 Del	1.74	Match
EGFRv3 25	L861Q	0.94	Discrepant
EGFRv3 28	L861Q	0.94	Match
EGFRv3 13	Exon 19 Del	0.86	Discrepant
EGFRv3 16	WT	0.64	Match
Mean PCR DNA input (ng)	Match:	54.35	**Concordance**
	Discrepant:	48.03	**39.13%**

Table A2.1.4: Anonymised Pathology samples tested using the GeneFirst EGFR PCR assay version 4. Positive concordance value (%) shown in bottom right cell. WT = Wild-type.

Study ID	Mutations	DNA Loading (ng)	Concordance with Therascreen
EGFRv4 02	T790M	78.54	Discrepant
EGFRv4 14	Exon 19 Del	56.96	Discrepant
EGFRv4 04	WT	35.60	Match
EGFRv4 06	L858R	33.60	Discrepant
EGFRv4 10	Exon 19 Del	32.21	Match
EGFRv4 13	L858R	31.60	Discrepant
EGFRv4 15	WT	30.60	Match
EGFRv4 05	T790M	17.46	Discrepant
EGFRv4 12	Exon 19 Del	16.80	Match
EGFRv4 01	Exon 19 Del	13.40	Discrepant
EGFRv4 11	Exon 19 Del	11.20	Discrepant
EGFRv4 07	L858R	2.56	Match
EGFRv4 03	Exon 19 Del	1.73	Discrepant
EGFRv4 08	L858R	1.28	Match
EGFRv4 09	L858R	0.51	Match
Mean PCR DNA input (ng)	Match:	**17.17**	**Concordance**
	Discrepant:	**30.48**	**46.67%**

Table A2.1.5 Anonymised Pathology samples tested using the GeneFirst EGFR PCR assay version 5. Positive concordance value (%) shown in bottom right cell. 12 samples contained DNA at a concentration below the lower threshold of the Qubit instrument, indicated with "Unknown (low)" in the DNA loading field. WT = Wild-type.

Study ID	Mutations	DNA loading (ng)	Concordance with Therascreen
EGFRv5 09	Exon 19 Del	30.00	Match
EGFRv5 01	Exon 20 Ins	30.00	Discrepant
EGFRv5 02	L861Q and G719X	30.00	Discrepant
EGFRv5 03	Exon 19 Del	30.00	Discrepant
EGFRv5 04	T790M	30.00	Discrepant
EGFRv5 05	L861Q	30.00	Discrepant
EGFRv5 06	L858R	30.00	Discrepant
EGFRv5 07	T790M	30.00	Discrepant
EGFRv5 08	S768I	30.00	Discrepant
EGFRv5 18	Exon 19 Del	15.00	Match
EGFRv5 10	Exon 20 Ins	15.00	Discrepant
EGFRv5 11	L861Q and G719X	15.00	Discrepant
EGFRv5 12	Exon 19 Del	15.00	Discrepant
EGFRv5 13	T790M	15.00	Discrepant
EGFRv5 14	L861Q	15.00	Discrepant
EGFRv5 15	L858R	15.00	Discrepant
EGFRv5 16	T790M	15.00	Discrepant
EGFRv5 17	S768I	15.00	Discrepant
EGFRv5 21	WT	Unknown (low)	Match
EGFRv5 22	WT	Unknown (low)	Match
EGFRv5 23	WT	Unknown (low)	Match
EGFRv5 24	WT	Unknown (low)	Match
EGFRv5 29	WT	Unknown (low)	Match
EGFRv5 30	WT	Unknown (low)	Match
EGFRv5 19	WT	Unknown (low)	Discrepant
EGFRv5 20	Del	Unknown (low)	Discrepant
EGFRv5 25	WT	Unknown (low)	Discrepant
EGFRv5 26	WT	Unknown (low)	Discrepant
EGFRv5 27	WT	Unknown (low)	Discrepant
EGFRv5 28	WT	Unknown (low)	Discrepant
Mean PCR DNA input (ng)	Match:	**22.50**	**Concordance**
	Discrepant:	**22.50**	**26.67%**

Table A2.1.6 Experiments running no template controls in the GF EGFR assay version 3. Control designations: 'NF H_2O' = Nuclease Free water control, tested in master mix A, B, C or D. 'MM only' = master mix only, (no H_2O added) master mix A, B, C or D. All no template controls (H_2O) gave a signal in the ROX channel. The only reaction to not give a ROX signal was one of the 'MM only' controls in master mix D. No signals were detected in the VIC and FAM channels.

		ROX		PCR VALIDATION		
Sample Name	PCR Master Mix	Ct	Ct Mean	Control Pos/ Neg	Assay Pass/ Fail	Comments
NF H_20	A	30.762	31.899	POS	FAIL	ROX positive
NF H_20	A	33.036	31.899	POS	FAIL	ROX positive
MM only	A	28.562	28.583	POS	FAIL	ROX positive
MM only	A	28.605	28.583	POS	FAIL	ROX positive
NF H_20	B	30.157	29.957	POS	FAIL	ROX positive
NF H_20	B	29.756	29.957	POS	FAIL	ROX positive
MM only	B	29.436	29.213	POS	FAIL	ROX positive
MM only	B	28.991	29.213	POS	FAIL	ROX positive
NF H_20	C	27.731	27.441	POS	FAIL	ROX positive
NF H_20	C	27.150	27.441	POS	FAIL	ROX positive
MM only	C	29.833	28.146	POS	FAIL	ROX positive
MM only	C	26.459	28.146	POS	FAIL	ROX positive
NF H_20	D	30.425	29.721	POS	FAIL	ROX positive
NF H_20	D	29.016	29.721	POS	FAIL	ROX positive
MM only	D	Undetermined	27.518	NEG	PASS	
MM only	D	27.518	27.518	POS	FAIL	ROX positive

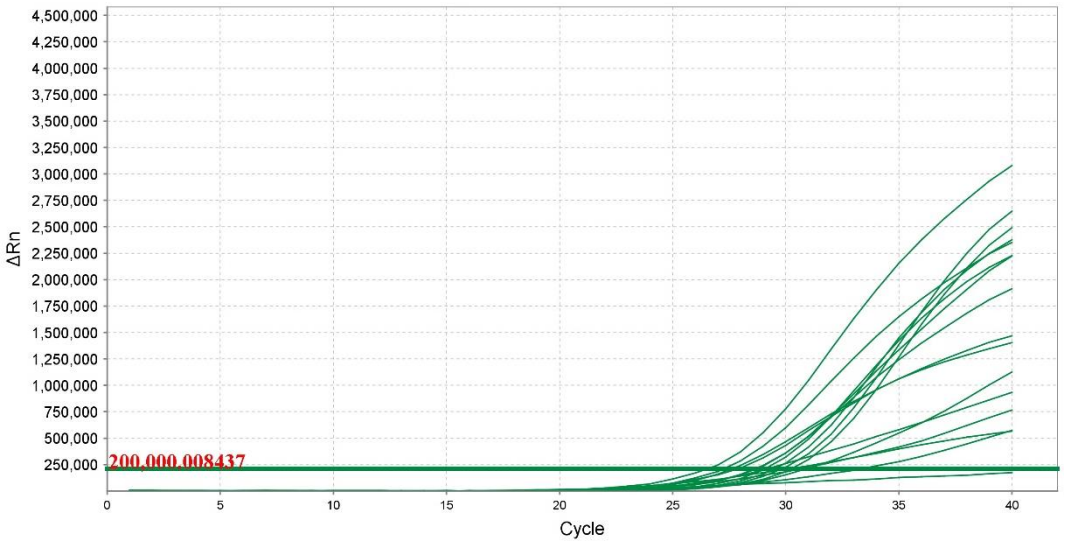

Figure A2.1 Amplification plot showing the ROX channel from the no template control experiment with GF EGFR assay version 3 (sample data shown in table A2.1.6). The plot shows a number of fluorescence signals being generated by no template controls. No signals were detected in the VIC and FAM channels.

A2.2 GeneFirst KRAS raw data versions 1-5

Table A2.2.1: Anonymised Pathology samples tested using the GeneFirst KRAS PCR assay version 1. Positive concordance value (%) shown in bottom right cell. WT = Wild-type.

Study ID	Mutations	DNA Loading (ng)	Concordance with CAST
KRASv1 01	G12A	30.00	Match
KRASv1 02	G12D	30.00	Match
KRASv1 03	WT	30.00	Match
KRASv1 04	G12A	30.00	Match
KRASv1 05	WT	30.00	Match
KRASv1 06	G12C	30.00	Match
KRASv1 07	G12V	30.00	Discrepant
KRASv1 08	G12A	15.00	Discrepant
KRASv1 09	G12D	15.00	Match
KRASv1 10	G12A	15.00	Match
KRASv1 11	WT	15.00	Match
KRASv1 16	G12C	15.00	Match
KRASv1 17	G12V	15.00	Match
KRASv1 18	WT	15.00	Match
KRASv1 12	G12A	7.50	Match
KRASv1 13	G12D	7.50	Match
KRASv1 14	G12A	7.50	Match
KRASv1 15	WT	7.50	Match
KRASv1 19	G12C	7.50	Match
KRASv1 20	G12V	7.50	Match
KRASv1 21	WT	7.50	Match
Mean PCR DNA input (ng)	Match:	**16.97**	**Concordance**
	Discrepant:	**22.50**	**90.48%**

Table A2.2.2: Anonymised Pathology samples tested using the GeneFirst KRAS PCR assay version 2. Positive concordance value (%) shown in bottom right cell. WT = Wild-type.

Study ID	Mutations	DNA Loading (ng)	Concordance with CAST
KRASv2 06	G12C	185.20	Match
KRASv2 07	G12C	185.20	Discrepant
KRASv2 10	G12V	111.31	Match
KRASv2 11	G12V	111.31	Match
KRASv2 08	G12A	86.53	Discrepant
KRASv2 09	G12A	86.53	Match
KRASv2 25	WT	85.93	Match
KRASv2 26	WT	84.33	Match
KRASv2 17	WT	63.62	Match
KRASv2 22	WT	61.97	Match
KRASv2 01	WT	61.41	Match
KRASv2 02	WT	61.41	Match
KRASv2 20	WT	59.72	Match
KRASv2 24	WT	52.89	Match
KRASv2 03	G12A	52.33	Discrepant
KRASv2 04	G12A	52.33	Discrepant
KRASv2 18	G12D	50.78	Discrepant
KRASv2 16	WT	47.31	Match
KRASv2 15	WT	35.72	Match
KRASv2 23	G12D	35.14	Discrepant
KRASv2 33	G12D	33.20	Discrepant
KRASv2 27	G12V	32.60	Match
KRASv2 30	G12V	31.42	Match
KRASv2 39	G12V	31.40	Match
KRASv2 38	WT	30.00	Match
KRASv2 32	G12V	29.04	Match
KRASv2 28	G12D	27.00	Discrepant
KRASv2 43	G12D	27.00	Match
KRASv2 31	G12D	25.00	Match
KRASv2 19	G12D	22.15	Match
KRASv2 21	G12V	21.90	Discrepant
KRASv2 29	G12D	21.00	Discrepant
KRASv2 05	G12D	8.18	Discrepant
KRASv2 40	G12V	3.00	Match
KRASv2 44	G12D	3.00	Match
KRASv2 41	G12V	1.50	Discrepant
KRASv2 45	G12D	1.50	Match
KRASv2 42	G12V	0.60	Match
KRASv2 46	G12D	0.60	Match
Mean PCR DNA input (ng)	Match:	**49.85**	**Concordance**
	Discrepant:	**47.92**	**69.23%**

Table A2.2.3 Anonymised Pathology samples tested using the GeneFirst KRAS PCR assay version 3. Positive concordance value (%) shown in bottom right cell. WT = Wild-type.

Study ID	Mutations	DNA Loading (ng)	Concordance with CAST
KRASv3 09	G12V	21.00	Match
KRASv3 07	G13D	20.80	Match
KRASv3 08	G12D	18.70	Discrepant
KRASv3 15	G12D	15.00	Discrepant
KRASv3 16	G12V	15.00	Match
KRASv3 17	WT	15.00	Match
KRASv3 18	G13D	15.00	Match
KRASv3 19	WT	15.00	Match
KRASv3 20	WT	15.00	Match
KRASv3 21	WT	15.00	Match
KRASv3 22	WT	15.00	Match
KRASv3 23	WT	15.00	Match
KRASv3 24	G12V	15.00	Match
KRASv3 25	WT	15.00	Match
KRASv3 01	G12A	6.84	Match
KRASv3 04	G12A	4.52	Match
KRASv3 03	G12C	4.32	Match
KRASv3 06	G12D	2.36	Discrepant
KRASv3 02	G12D	2.14	Discrepant
Mean PCR DNA input (ng)	Match:	**13.83**	**Concordance**
	Discrepant:	**9.55**	**78.95%**

Table A2.2.4 Anonymised Pathology samples tested using the GeneFirst KRAS PCR assay version 4. Positive concordance value (%) shown in bottom right cell. WT = Wild-type.

Study ID	Mutations	DNA Loading (ng)	Concordance with CAST
KRASv4 01	G12D	24.40	Discrepant
KRASv4 10	G12V	16.00	Match
KRASv4 07	WT	11.44	Discrepant
KRASv4 03	WT	10.12	Discrepant
KRASv4 04	WT	9.32	Discrepant
KRASv4 08	G12A	8.72	Discrepant
KRASv4 09	G12C	7.16	Match
KRASv4 05	WT	5.56	Discrepant
KRASv4 02	WT	5.44	Discrepant
KRASv4 14	G12V	4.24	Match
KRASv4 15	G12D	4.16	Discrepant
KRASv4 13	WT	3.26	Discrepant
KRASv4 12	WT	1.00	Match
KRASv4 06	WT	0.71	Discrepant
KRASv4 11	WT	0.39	Discrepant
Mean PCR DNA input (ng)	Match:	**7.10**	**Concordance**
	Discrepant:	**7.59**	**26.67%**

Table A2.2.5 Anonymised Pathology samples tested using the GeneFirst KRAS PCR assay version 5. Positive concordance value (%) shown in bottom right cell. WT = Wild-type.

Study ID	Mutations	DNA Loading (ng)	Concordance with CAST
KRASv5 01	WT	30.00	Match
KRASv5 02	WT	30.00	Match
KRASv5 03	G12V	30.00	Match
KRASv5 04	G12C	30.00	Match
KRASv5 05	G12C	30.00	Match
KRASv5 06	G12D	30.00	**Discrepant**
KRASv5 07	G12D	30.00	Match
KRASv5 08	G12A	30.00	**Discrepant**
KRASv5 09	WT	30.00	**Discrepant**
KRASv5 10	WT	15.00	Match
KRASv5 11	WT	15.00	Match
KRASv5 12	G12V	15.00	**Discrepant**
KRASv5 13	G12C	15.00	Match
KRASv5 14	G12C	15.00	Match
KRASv5 15	G12D	15.00	Match
KRASv5 16	G12D	15.00	Match
KRASv5 17	G12A	15.00	**Discrepant**
KRASv5 18	WT	15.00	**Discrepant**
Mean PCR DNA input (ng)	Match:	**22.50**	**Concordance**
	Discrepant:	**22.50**	**66.67%**

Table A2.2.6 Experiments running no template controls in the GF KRAS assay version 3. Control designations: 'NF H_2O' = Nuclease Free water control, tested in master mix A or B. 'MM only' = master mix only, (no H_2O added) master mix A or B. All controls (H_2O and MM only) gave a signal in the ROX channel. Additionally the MM only control in master mix B gave a signal in the FAM channel.

		FAM		ROX		PCR VALIDATION		
Sample Name	PCR Master Mix	Ct	Ct Mean	Ct	Ct Mean	Control Pos/ Neg	Assay Pass/ Fail	Comments
NF H_20	A	Undetermined		15.374	15.345	POS	FAIL	ROX positive
NF H_20	A	Undetermined		15.316	15.345	POS	FAIL	ROX positive
MM only	A	Undetermined		14.606	14.730	POS	FAIL	ROX positive
MM only	A	Undetermined		14.854	14.730	POS	FAIL	ROX positive
NF H_20	B	Undetermined		17.624	17.790	POS	FAIL	ROX positive
NF H_20	B	Undetermined		17.956	17.790	POS	FAIL	ROX positive
MM only	B	24.285	24.697	17.450	17.476	POS	FAIL	ROX positive
MM only	B	25.108	24.697	17.502	17.476	POS	FAIL	ROX positive

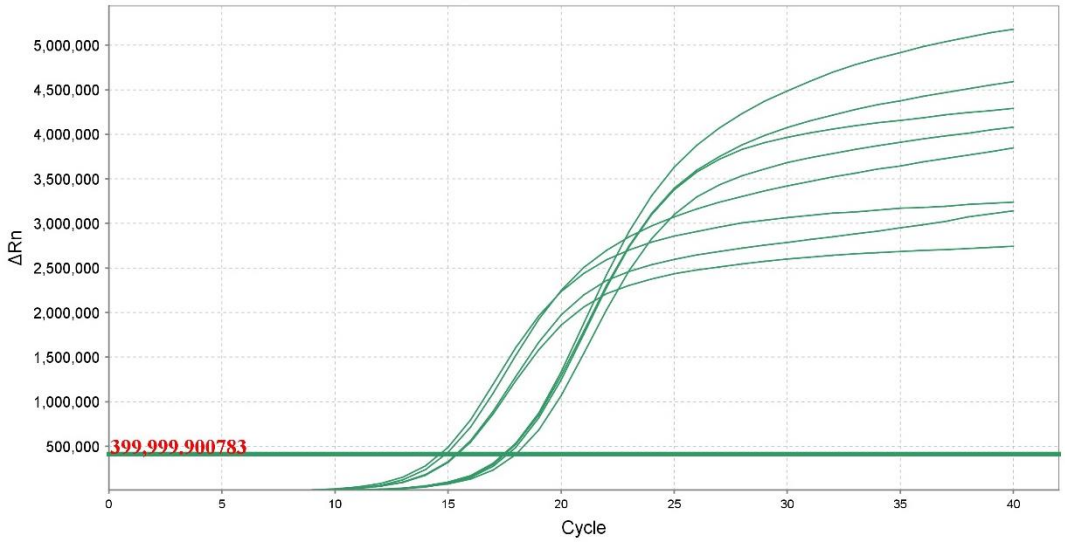

Figure A2.2.1 Amplification plot showing the ROX channel from the no template control experiment with GF KRAS assay version 3 (sample data shown in table A2.2.6). The plot shows a number of fluorescence signals being generated by no template controls. No signals were detected in the VIC channel.

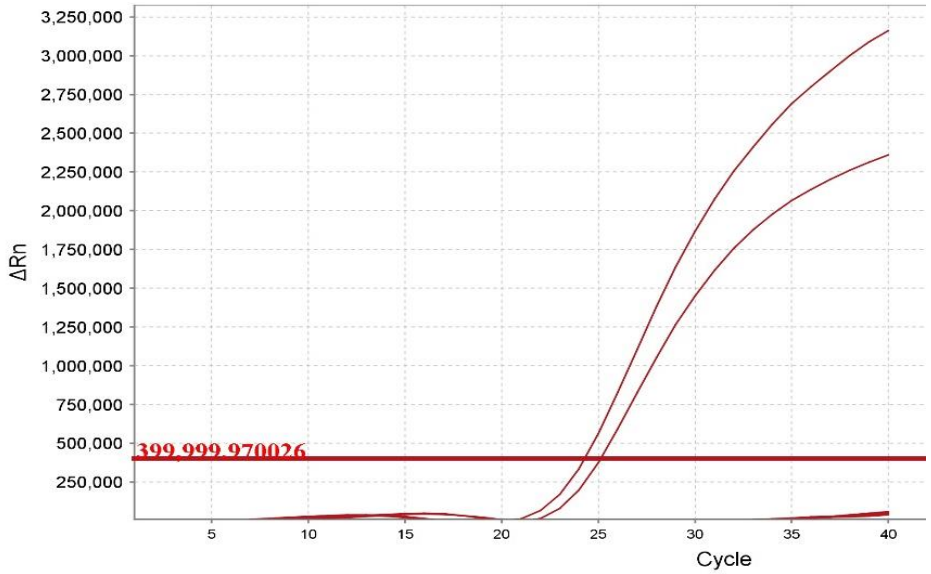

Figure A2.2.2 Amplification plot showing the FAM channel from the no template control experiment with GF KRAS assay version 3 (sample data shown in table A2.2.6). The plot shows the two fluorescence signals generated by the master mix only control in reaction B. No signals were detected in the VIC channel.

A2.3 NGS Supplemental Data

Table A2.3 All NGS hotspot data detected in tissue DNA samples that were tested using the GeneFirst EGFR or KRAS assays. Rows containing *EGFR* and *KRAS* hotspots are highlighted in blue. SNP = Single Nucleotide Polymorphism.

Study ID	Frequency	Quality	Type	Allele Name	AA change	Gene ID	Coverage	Allele Coverage	Strand Bias
EGFRv2 43	2	7.28	SNP	COSM232263	p.R612Q	ERBB4	152	3	0.6714
EGFRv2 43	66.7	1974.18	SNP	COSM125370	p.E545K	PIK3CA	255	170	0.5217
EGFRv2 43	66.7	1974.18	SNP	COSM763	p.E545K	PIK3CA	255	170	0.5217
EGFRv2 43	3.6	7.92	SNP	COSM22973	p.R393*	FBXW7	56	2	0.5884
EGFRv2 43	2.8	12.52	SNP	COSM17570	p.P741L	EGFR	177	5	0.5527
EGFRv2 43	2	7.48	SNP	COSM33729	p.R603*	BRAF	252	5	0.7722
EGFRv2 43	41.3	774.99	SNP	COSM522	p.G12A	KRAS	213	88	0.5634
EGFRv2 43	3.4	7.40	SNP	COSM11183	p.R267W	TP53	119	4	0.684
EGFRv2 43	3.4	7.40	SNP	COSM179804	p.R267W	TP53	119	4	0.684
EGFRv2 43	1.9	6.39	SNP	COSM6932	p.G245S	TP53	312	6	0.6191
EGFRv2 43	1.9	6.39	SNP	COSM121037	p.G152S	TP53	312	6	0.6191
EGFRv2 43	1.9	6.39	SNP	COSM121036	p.G245S	TP53	312	6	0.6191
EGFRv2 43	1.9	6.39	SNP	COSM121035	p.G245S	TP53	312	6	0.6191
EGFRv2 43	2.9	10.97	SNP	COSM43939	p.T211I	TP53	139	4	0.5394
EGFRv2 43	2.3	8.60	SNP	COSM45440	p.A189A	TP53	133	3	0.7497
EGFRv2 43	2.6	11.23	SNP	COSM45627	p.I162I	TP53	229	6	0.5239
EGFRv2 43	5.5	13.23	SNP	COSM44492	p.W91*	TP53	55	3	0.5331
EGFRv2 44	2.8	12.34	SNP	COSM232263	p.R612Q	ERBB4	179	5	0.8275
EGFRv2 44	2.3	8.55	SNP	COSM28605	p.A871T	EGFR	133	3	0.5617
EGFRv2 44	6.7	9.81	DEL	COSM13047	p.V1578delV	NOTCH1	30	2	0.533
EGFRv2 44	25.8	401.81	SNP	COSM516	p.G12C	KRAS	229	59	0.5199
EGFRv2 44	2.9	11.86	SNP	COSM44737	p.E287G	TP53	140	4	0.7434
EGFRv2 44	16.8	122.24	SNP	COSM10859	p.V272L	TP53	137	23	0.5399
EGFRv2 44	21.8	388.92	SNP	COSM11517	p.Y126C	TP53	289	63	0.5352
EGFRv2 44	36.7	193.51	DEL	COSM20871	p.P281fs*6	STK11	139	51	0.5651
EGFRv2 44	36.7	193.51	DEL	COSM12924	p.P281fs*6	STK11	139	51	0.5651
EGFRv2 46	2.4	8.37	DEL	COSM13047	p.V1578delV	NOTCH1	82	2	0.5604
EGFRv2 46	1.7	6.62	DEL	COSM5823	p.T321fs*23	PTEN	287	5	0.6341
EGFRv2 46	1.7	6.62	DEL	COSM5801	p.N323fs*21	PTEN	287	5	0.6341
EGFRv2 46	3	11.83	SNP	COSM44194	p.A84V	TP53	135	4	0.5037
EGFRv2 46	2.1	9.06	SNP	COSM14065	p.V842I	ERBB2	328	7	0.526
EGFRv2 04	88.6	12667.50	SNP	COSM6224	p.L858R	EGFR	1996	1768	0.5073
EGFRv2 04	52.6	4657.79	SNP	COSM710	no record	MET	855	450	0.5017
KRASv2 05	1.9	6.16	SNP	COSM449	p.G464E	BRAF	215	4	0.7641
KRASv2 05	2.4	9.40	SNP	COSM13042	p.L1593P	NOTCH1	123	3	0.6971
KRASv2 05	1.8	6.60	SNP	COSM12772	p.L1574P	NOTCH1	114	2	0.5347
KRASv2 05	1.8	6.60	SNP	COSM24673	no record	NOTCH1	114	2	0.5347
KRASv2 05	3.5	9.21	SNP	COSM521	p.G12D	KRAS	260	9	0.7188
KRASv3 08	50.3	7922.22	SNP	COSM521	p.G12D	KRAS	1546	777	0.5096
KRASv2 03	33.7	5434.23	SNP	COSM760	p.E542K	PIK3CA	1999	674	0.5038
KRASv2 03	40.3	7238.55	SNP	COSM522	p.G12A	KRAS	2000	806	0.5094
KRASv2 07	39	6870.49	SNP	COSM760	p.E542K	PIK3CA	2000	780	0.5222
KRASv2 07	48	4827.25	SNP	COSM516	p.G12C	KRAS	1019	489	0.5117
KRASv2 07	78.2	1989.65	SNP	COSM10660	p.R273H	TP53	193	151	0.5143

Chapter Three – Optimisation and Validation of Cancer Gene Mutation Detection in tissue using TaqMan Array

3.1 INTRODUCTION

The advancement of molecular biology over recent decades has revolutionised diagnostic medicine by allowing rapid, accurate and cost effective analysis of genetic aspects of human disease pathology. Current developments in personalised medicine for the treatment of cancer would not be possible without both the expansion of our knowledge of actionable mutations and recent developments in molecular technology. Diagnosis rates for many cancers are also improving dramatically, resulting in many more patients being diagnosed and managed at an earlier stage.

However, the expansion of our knowledge and increased diagnosis rate has put both additional strain on the NHS pathology services and an increased demand for analysis of molecular targets. Molecular diagnostic laboratories are under increasing pressure to offer a broader range of molecular assays with a fast turnaround time. In order to meet this demand, NHS pathology services must continuously assess their current assays and investigate new technology that can maximise staff time efficiency, maintain standards of accuracy and precision, and potentially broaden the range of services the department can offer to their clinical colleagues.

In order to meet these requirements of a rapidly developing field, many current molecular assays have become increasingly complex and labour intensive to perform. Increasing complexity of assays leaves them vulnerable to human error and contamination, resulting in potential mistakes or misdiagnosis, and also resulting in assays needing to be repeated and thus a knock on affect to turnaround time and cost.

In the UHCW Pathology Department *EGFR*, *KRAS* and *BRAF* testing are performed on separate assays (Qiagen Therascreen for *EGFR*, an in-house Life Technologies castPCR for *KRAS* and *BRAF*). Technical details of these assays are described in the Methods section. These assays have both been clinically validated for diagnostic use, however they both require the use of multiple master mixes and a considerable amount of staff hands-on-time due to multiple replicates required per sample, and a high degree of manual dexterity due to the small RotaGene™ tubes (Therascreen) or microtiter plate wells (CastPCR) used to run the reactions.

The TaqMan Array (TA) is a PCR platform technology developed by Life Technologies (ThermoFisher). The array comes uses a 384 well plate format, with each set of 48 wells separated into 8 micro-fluidic channels. Each of the 48 wells in the channels contains a set of lyophilised primers and probes for one specific molecular target dispensed during manufacture of the array. Each well has a reaction volume of just 1 microlitre. A mixture of sample and master mix reagents are dispensed into a loading port at one end of the plate, which is then centrifuged to distribute the sample-reagent mixture across all 48 wells in that channel. The chemistry for these individual well assays is the same as castPCR (described previously). 48 specific molecular targets can be analysed per sample per run. This is a considerable improvement on other assays as in this case every target has its own well, whereas other assays often detect multiple mutants in the same well, but cannot distinguish between them (e.g. GeneFirst).

Once the sample and reagents are centrifuged into the TA wells, the plate can be sealed and run on a ViiA7 Thermocycler fitted with a compatible heating block and lid. As the plate is separated into channels of 48 wells, therefore 8 samples (typically 7 samples and 1 negative control) can be analysed on one plate, and the entire plate can be loaded in only 8 pipetting

actions. This dramatically reduces operator hands on time and the chances of contamination error.

Beyond potential diagnostic applications, TA format assays are widely used in a number of research settings, most notably for the analysis of microRNA (miRNA) samples. In the field of cancer there are a number of recent studies using TA to analyse miRNA samples to investigate a range of different cancers, including: esophageal cancer (Warnecke-Eberz *et al.*, 2016), gastric cancer (Qiu *et al.*, 2016; Zhu *et al.*, 2014), hepatocellular carcinoma (Zhuang *et al.*, 2015) and ovarian cancer (Zheng *et al.*, 2013). Outside cancer TA assays are utilised in a number of other disciplines including a range of microbiology applications (Muhammad *et al.*, 2015; Joshi & Arankalle, 2015; Aqil *et al.*, 2014), cardiology (Sepramaniam *et al.*, 2014) and psychiatry/ neurology (Wei *et al.*, 2015). The diversity of fields in which the TA has been applied strongly suggests this format of molecular assay is both robust and adaptable, so should mean it is well suited to the detection of *EGFR* and *KRAS* mutations in our clinical setting.

3.1.1 Aims

The Taqman Array assay has a number of considerable advantages over the two currently used assays in UHCW Pathology. It has the potential to combine several gene mutation assays in one plate, with reduced hands-on time and equivalent sensitivity. The aim of the study was to replace the current Qiagen Therascreen EGFR assay and KRAS castPCR assays with a single array, producing savings for the department in staff time and resources while providing patients with an improved molecular diagnostic service. The combined RAS, EGFR and BRAF assay is known as the REB array.

- To optimise the sample loading specifications for analysis of *EGFR*, *KRAS*, *NRAS* and *BRAF* from extracted FFPE tissue samples.

- To validate the performance of the REB array assay against the currently used Qiagen Therascreen EGFR assay and the Life Technologies castPCR assay. Results to be verified by Ion Torrent NGS analysis.

- If the TA assay gives comparable or improved results over currently used assays: the TA will be implemented by UHCW Pathology as their new service method for detection of *EGFR*, *KRAS*, *NRAS* and *BRAF* mutations.

3.2 METHODS

3.2.1 Equipment

Bench top centrifuge compatible with 1.5ml micro-centrifuge tubes (up to 15500g)

Bench top centrifuge compatible with 96 well PCR plates (up to 1600g) and Taqman arrays

Bench top centrifuge compatible with Ion Torrent sequencing chips (314, 316 and 318)

Heat block capable of heating 1.5ml tubes to 70°C

Promega Maxwell Automated Extraction Instrument

Laminar Flow safety cabinet with UV decontamination feature

Qubit 2.0 Nucleic Acid Analyser

Life Technologies ViiA 7 Dx Real Time PCR instrument with Fast 96 well block

Life Technologies ViiA 7 384 well block

Life Technologies TaqMan Array card staker (for TA card sealing)

Life Technologies Ion Chef instrument (see figure 2.2.8.3.2)

Life Technologies Ion Torrent PGM instrument (see figure 2.2.8.3.2)

Millipore water purification system

3.2.2 Kits and Reagents

Qubit dsDNA High Sensitivity reagent kit

Promega Maxwell FFPE DNA Kit

Promega Maxwell Circulating DNA Kit (prototype)

Nuclease free DEPC-treated water

TaqMan Universal PCR Master Mix

Life Technologies Ion AmpliSeq™ Colon and Lung Panel

Life Technologies Ion Express Barcode 1-16 Kit

Life Technologies Ion Express Barcode 17- 32 Kit

Life Technologies Ion Torrent PGM Sequencing 200 Kit

Life Technologies Ion AmpliSeq™ Library Kit 2.0

Life Technologies Ion PGM IC 200 Kit

Life Technologies Oncomine™ Solid Tumour DNA Kit

3.2.3 Design of REB Arrays

The plate design is shown in table 3.2.3. In discussion with my supervisor Prof Cree, the COSMIC database was interrogated to identify mutations that were most commonly reported, and the resultant list was used to compile the plate playout. CastPCR assays for each of these were identified from the Life technologies website (http://www.lifetechnologies.com/it/en/home/life-science/pcr/real-time-pcr/real-time-pcr-assays/taqman-mutation-detection-assays/somatic-mutation-real-time-pcr.html) and used to populate a spreadsheet which was used to manufacture the array cards. TaqMan® Mutation Detection Assays for the detection of mutant alleles must be run in parallel with corresponding wild-type allele assays. This acts as a positive control for the wild-type DNA. For each gene, a reference assay was therefore required.

Table 3.2.3 Reference (suffix _rf) and Mutant (suffix _mu) assays included in the TA assay plate lay out.

Column	Row	TA Ref	HGVS	Row	TA Ref	HGVS
1	A	EGFR_rf	-	B	KRAS_522_mu	KRAS c.35G>C p.Gly12Ala
2	A	EGFR_ex19dels_mu	EGFR c.2238_2252del15 p.L747_T751delLREAT	B	KRAS_518_mu	KRAS c.34G>C p.Gly12Arg
3	A	EGFR_6224_mu	EGFR c.2573T>G p.Leu858Arg; EGFR c.2572_2573CT>AG p.Leu858Arg	B	KRAS_19404_mu	KRAS c.436G>A p.Ala146Thr
4	A	EGFR_6240_mu	EGFR c.2369C>T p.Thr790Met	B	KRAS_527_mu	KRAS c.37G>T p.Gly13Cys
5	A	EGFR_6213_mu	EGFR c.2582T>A p.Leu861Gln	B	KRAS_554_mu	KRAS c.183A>C p.Gln61H
6	A	EGFR_6241_mu	c.2303G>T p.Ser768Ile	B	KRAS_553_mu	KRAS c.182A>T p.Gln61Leu
7	A	EGFR_6239_mu	EGFR c.2156G>C Gly719Ala	B	KRAS_529_mu	KRAS c.37G>C p.Gly13Arg
8	A	EGFR_6252_mu	EGFR c.2155G>A p.Gly719Ser	B	NRAS_rf	-
9	A	EGFR_6253_mu	EGFR c.2155G>T P.Gly719Cys	B	NRAS_569_mu	NRAS c.37G>C p.Gly13Arg
10	A	EGFR_12988_mu	EGFR c.2125G>A p.Glu709Lys	B	NRAS_562_mu	NRAS c.34G>T p.Gly12Cys
11	A	EGFR_13427_mu	EGFR c.2126A>C p.Glu709Ala	B	NRAS_563_mu	NRAS c.34G>A p.Gly12Ser
12	A	EGFR_13428_mu	EGFR c.2311_2312insGCGTGGACA	B	NRAS_573_mu	NRAS c.38G>A p.Gly13Asp
13	A	EGFR_12381_mu	EGFR c.2319_2320ins9	B	NRAS_583_mu	NRAS c.181_182CA>TT
14	A	BRAF_rf	-	B	NRAS_564_mu	NRAS c.35G>A p.Gly12Asp
15	A	BRAF_476_mu	BRAF c.1799T>A p.Val600Glu	B	NRAS_580_mu	NRAS c.181C>A p.Gln61Lys
16	A	BRAF_475_mu	BRAF c.1798_1799GT>AA p.Val600Glu	B	NRAS_584_mu	NRAS c.181_182CA>AG
17	A	BRAF_474_mu	BRAF c.1798_1799GT>AG p.Val600Arg	B	NRAS_574_mu	NRAS c.38G>T p.Gly13Val
18	A	BRAF_473_mu	BRAF c.1799_1800TG>AT p.Val600Asp	B	NRAS_566_mu	NRAS c.35G>T p.Gly12Val
19	A	KRAS_rf	-	B	NRAS_585_mu	NRAS c.183A>T p.Gln61His
20	A	KRAS_521_mu	KRAS c.35G>A p.Gly12Asp	B	NRAS_586_mu	NRAS c.183A>C p.Gln61His
21	A	KRAS_520_mu	KRAS c.35G>T p.Gly12Val	B	NRAS_565_mu	NRAS c.35G>C p.Gly12Ala
22	A	KRAS_532_mu	KRAS c.38G>A p.Gly13Asp	B	NRAS_570_mu	NRAS c.37G>T p.Gly13Cys
23	A	KRAS_516_mu	KRAS c.34G>T p.Gly12Cys	B	NRAS_571_mu	NRAS c.37G>A p.Gly13Ser
24	A	KRAS_517_mu	KRAS c.34G>A p.Gly12Ser	B	NRAS_575_mu	NRAS c.38G>C p.Gly13Ala

Each well on the 384 well REB Array contains a separate assay, allowing detection of a large number of targets, without multiplexing. The card is divided along the rows into 8 separate channels, each containing 48 wells housing separate castPCR assays (four reference wild-type assays, 44 mutant assays) into which a single sample is loaded with a single pipetting step. This elimination of individual well pipetting dramatically reduces the risk of contamination or operator error. As each port uses 48 wells, and the individual assays are very reliable (CoV < 2% in previous studies (Glaysher *et al.*, 2009)), a single well is sufficient for each assay. It is however, good practice to run control plates at regular intervals, particularly with new plate or reagent batches. If duplicates or triplicates for individual samples or controls are required, two or three ports respectively must be used.

3.2.4 Patients and Samples

96 patients were retrospectively identified with known mutational status. Samples were either extracted DNA from remaining formalin-fixed, paraffin-embedded FFPE blocks, or used stored DNA from the previous diagnostic PCR assays to validate the Taqman/ REB array, the characteristics according to the BRISQ (Biospecimen reporting for improved study quality) guidance (Moore *et al.*, 2011). Due to its retrospective nature for laboratory developed test validation, UK Health Research Authority Ethics Committee approval was not required.

FFPE blocks were identified for a series of patients with NSCLC (n = 42), colorectal cancer (n = 26), and melanoma (n = 28) for which mutational analysis previously been performed using Therascreen or castPCR. It should be noted that these were not consecutive or randomly selected patients, but were instead chosen reflect the range of mutations seen in our laboratory, for the purpose of clinical validation of the REB array from samples with sufficient remaining tissue surplus to diagnostic requirements. Where it was necessary to obtain further DNA, the same block was used when this could be identified from the pathology report.

3.2.5 DNA Extraction

Areas of high cellularity for neoplastic cells within the tumour were identified by a histopathologist and marked on a slide. These were matched with the corresponding block and case. Samples from the areas marked were punched out using a 1 mm diameter skin punch, as previously described (Bolton *et al.*, 2015). The samples were then placed in the

vial of a Maxwell DNA extraction robot (Promega, Southampton, UK) and DNA extracted according to the manufacturer's instructions. Stored or newly extracted DNA was checked for content and quality using a Nanodrop or Qubit instrument (both from Thermo Fisher Scientific) according to the manufacturer's instructions.

3.2.6 TaqMan Arrays

Individual samples were diluted with MM and dH_2O to give a volume of 100 µL. The reaction mixtures were prepared at varying DNA concentrations, depending on which phase of the validation was being performed (see Results). After mixing, 100 µL of each sample extract was added to each port of a REB array. The array plates were sealed, spun and placed in a previously calibrated ViiA7 PCR machine (ThermoFisher). Standard PCR conditions were used as follows: 50°C for 2 min, followed by 95°C for 1 min, then 40 cycles of 95°C for 15 sec, and 60°C for 1 min. Each run took 118 minutes to complete.

For efficient use of the REB arrays, samples must be grouped into batches of 7 (to fill the 8 channel card, 7 samples + 1 negative control). Smaller batch sizes require empty channels to be loaded with master mix before running, to avoid inaccurate data analysis by the ViiA7 software.

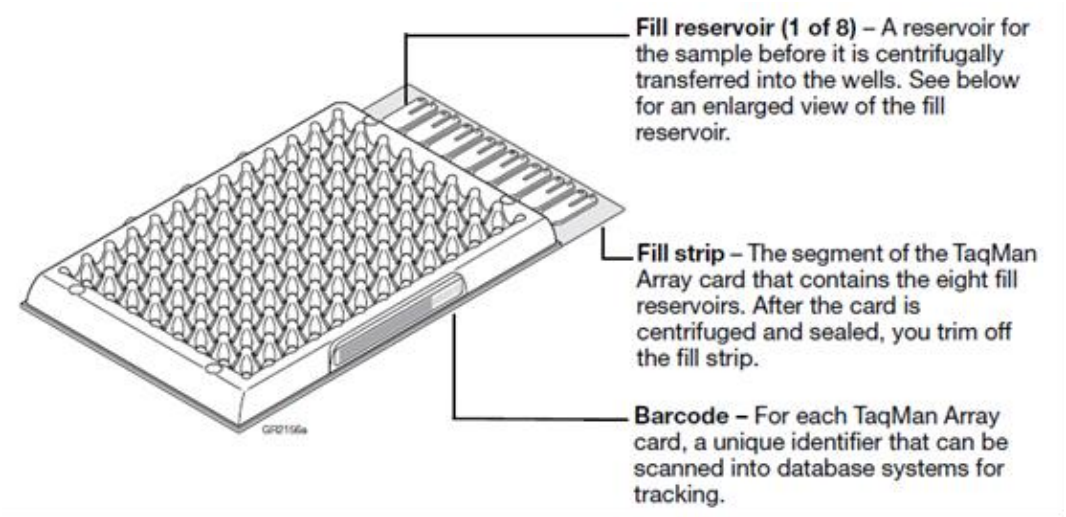

Figure 3.2.6 Diagram of the TaqMan Array card, from kit handbook (Thermofisher). 48 well channels are loaded using one of the eight fill reservoirs located in the fill strip at the open end of the card. The fill strip is removed after the card is sealed.

3.2.7 Data Analysis

The ViiA7 software produces a spreadsheet of Ct values for each well, clearly identified by sample number. The results from the spreadsheets were then collected into a further summary spreadsheet in Excel (Microsoft). The following exclusions were applied:

- Samples with insufficient DNA content (<50 ng total card input) were excluded from the analysis.
- Samples where one or more controls failed were excluded from the analysis as failed samples.
- Samples giving positive Ct values for targets on genes not present in the comparator assays were excluded from analysis as the significance of these results could not be determined.

Cases with discrepant results were sequenced on the Ion Torrent PGM using the OncoNetwork 22 gene panel (Tops *et al.*, 2015). Descriptive statistics, including sensitivity and specificity were produced using Excel.

3.3 RESULTS

3.3.1 Setting Analysis Thresholds

The first priority of the optimisation process was to determine the cycle thresholds for classifying samples as positive or negative. The manufacturer's guidance suggested a cut off Ct of 32 for positive PCR signals, however I considered it important to assess the performance of the assay with our sample type and select cut offs appropriate to our setting.

36 samples previously analysed FFPE DNA samples were obtained from the Pathology Department. The samples were loaded and run on the TaqMan Array cards at a range of DNA concentrations in order to assess the performance of the assay. Samples with high DNA concentrations or large remaining volumes were loaded at higher DNA inputs (where possible). Samples with low DNA concentrations or low remaining volumes were not diluted.

Table 3.2.1.1 shows the performance of the TaqMan Array control assays with the panel of 36 samples at varying concentrations. Ct values within these data varied greatly from 25.76 to 39.58. Using the manufacturer's suggested cut off resulted in 28/ 36 (77.8%) failing to generate a 'positive Ct' in all control assays. This suggested that the manufacturer's guidance was not necessarily suited to our sample types.

Table 3.3.1.1 TaqMan Array control assay performance with a selection of samples at varying DNA concentrations. Ranked and colour scaled in order of gDNA input, lowest to highest. Ct values considered positive are below 32, highlighted in green/ blue. Assays failing for generate a Ct are represented as 'Undetermined' (highlighted red/ pink).

Study Sample No	gDNA input (ng, total)	EGFR_Rf Ct	KRAS_Rf Ct	NRAS_Rf Ct	BRAF_Rf Ct
1	18.0	31.55	29.99	31.06	31.22
8	30.0	34.85	32.51	32.80	33.35
32	31.3	28.66	28.73	30.59	30.11
3	47.8	33.36	29.88	31.28	31.12
5	60.0	34.82	38.67	37.70	Undetermined
31	62.5	27.64	27.47	29.58	27.64
28	75.0	33.29	30.82	33.70	32.30
36	100.0	33.65	30.73	31.17	32.03
4	100.2	32.63	29.81	31.36	30.96
30	125.0	28.92	27.70	29.83	29.21
27	150.0	33.43	29.14	31.15	30.54
35	239.0	33.40	27.32	30.80	31.86
29	250.0	28.64	25.76	28.87	27.94
10	250.0	31.57	28.97	30.70	31.12
7	250.0	32.41	31.27	32.51	32.51
19	250.0	33.60	33.08	30.34	31.59
25	250.0	34.68	36.43	31.65	32.68
24	250.0	34.92	36.94	30.65	33.79
15	250.0	35.02	37.59	31.80	33.15
16	250.0	35.11	37.66	31.71	32.69
23	250.0	35.27	Undetermined	31.11	33.55
22	250.0	35.33	38.31	30.85	32.74
17	250.0	36.03	Undetermined	31.66	33.77
21	250.0	36.80	Undetermined	32.27	33.84
12	250.0	37.08	Undetermined	31.59	34.99
18	250.0	38.50	39.58	33.23	37.17
14	250.0	39.45	Undetermined	33.17	37.76
13	250.0	Undetermined	Undetermined	35.50	Undetermined
20	250.0	Undetermined	Undetermined	Undetermined	Undetermined
33	300.0	33.69	28.68	30.31	30.63
34	300.0	Undetermined	31.94	33.21	33.45
2	302.8	36.63	33.24	32.96	32.35
26	450.0	30.77	27.22	28.64	29.97
9	500.0	29.18	27.35	29.01	29.06
11	500.0	32.11	28.99	31.12	31.66
6	500.0	33.08	29.72	32.17	31.58

The performance of the mutant target assays is shown in table 3.2.1.2. In total 31 mutant assays gave a detectable signal. Using the manufacturer's guidance cut off value of Ct32 for positive results, only 2/ 31 (6.5%) were determined as positive. Again this suggests the guidance values are not suited to our sample type.

The results from this initial data set were reanalysed using higher cut off values. Table 3.3.1.3 shows the reanalysed data from the control assays using a higher cut off of Ct35. This adjusted parameter resulted in a reduction in failure rate to 16/ 36 (44.4%), and reduction of 33.4% over the first analysis.

Table 3.3.1.2 TaqMan Array mutant target assay performance with a selection of samples at varying DNA concentrations. Ranked and colour scaled in order of gDNA input, lowest to highest. Ct values considered positive are below 32, highlighted in green/ blue.

Sample No	DNA input (ng, total)	Target Name	Ct Mean
1	18.0	BRAF_475_mu	37.00
1	18.0	BRAF_476_mu	38.23
1	18.0	NRAS_564_mu	38.10
9	30.0	BRAF_477_mu	33.99
32	31.3	NRAS_573_mu	39.86
3	47.8	NRAS_584_mu	34.24
8	60.0	BRAF_476_mu	36.18
31	62.5	BRAF_476_mu	39.48
31	62.5	EGFR_6240_mu	39.32
28	75.0	KRAS_19404_mu	35.31
36	100.0	KRAS_553_mu	31.44
4	100.2	KRAS_553_mu	32.87
27	150.0	KRAS_19404_mu	34.08
35	239.0	NRAS_584_mu	33.42
7	250.0	BRAF_476_mu	35.53
10	250.0	KRAS_520_mu	34.62
12	250.0	NRAS_580_mu	31.93
14	250.0	EGFR_6240_mu	39.17
16	250.0	BRAF_474_mu	32.58
16	250.0	NRAS_564_mu	32.56
19	250.0	BRAF_476_mu	34.66
20	250.0	EGFR_6253_mu	36.26
20	250.0	NRAS_565_mu	37.77
21	250.0	BRAF_476_mu	36.25
24	250.0	BRAF_476_mu	37.32
33	300.0	BRAF_476_mu	36.38
2	302.8	NRAS_583_mu	38.95
26	450.0	KRAS_19404_mu	32.99
6	500.0	BRAF_476_mu	34.02
11	500.0	EGFR_6240_mu	36.99
11	500.0	KRAS_520_mu	32.99

Table 3.3.1.3 TaqMan Array control assay performance with a selection of samples at varying DNA concentrations. Ranked and colour scaled in order of gDNA input, lowest to highest. Ct values considered positive are below 35, highlighted in green/ blue. Assays failing for generate a Ct are represented as 'Undetermined' (highlighted red/ pink).

Study Sample No	gDNA input (ng, total)	EGFR_Rf Ct	KRAS_Rf Ct	NRAS_Rf Ct	BRAF_Rf Ct
1	18.0	31.55	29.99	31.06	31.22
8	30.0	34.85	32.51	32.80	33.35
32	31.3	28.66	28.73	30.59	30.11
3	47.8	33.36	29.88	31.28	31.12
5	60.0	34.82	38.67	37.70	Undetermined
31	62.5	27.64	27.47	29.58	27.64
28	75.0	33.29	30.82	33.70	32.30
36	100.0	33.65	30.73	31.17	32.03
4	100.2	32.63	29.81	31.36	30.96
30	125.0	28.92	27.70	29.83	29.21
27	150.0	33.43	29.14	31.15	30.54
35	239.0	33.40	27.32	30.80	31.86
29	250.0	28.64	25.76	28.87	27.94
10	250.0	31.57	28.97	30.70	31.12
7	250.0	32.41	31.27	32.51	32.51
19	250.0	33.60	33.08	30.34	31.59
25	250.0	34.68	36.43	31.65	32.68
24	250.0	34.92	36.94	30.65	33.79
15	250.0	35.02	37.59	31.80	33.15
16	250.0	35.11	37.66	31.71	32.69
23	250.0	35.27	Undetermined	31.11	33.55
22	250.0	35.33	38.31	30.85	32.74
17	250.0	36.03	Undetermined	31.66	33.77
21	250.0	36.80	Undetermined	32.27	33.84
12	250.0	37.08	Undetermined	31.59	34.99
18	250.0	38.50	39.58	33.23	37.17
14	250.0	39.45	Undetermined	33.17	37.76
13	250.0	Undetermined	Undetermined	35.50	Undetermined
20	250.0	Undetermined	Undetermined	Undetermined	Undetermined
33	300.0	33.69	28.68	30.31	30.63
34	300.0	Undetermined	31.94	33.21	33.45
2	302.8	36.63	33.24	32.96	32.35
26	450.0	30.77	27.22	28.64	29.97
9	500.0	29.18	27.35	29.01	29.06
11	500.0	32.11	28.99	31.12	31.66
6	500.0	33.08	29.72	32.17	31.58

Table 3.3.1.4 shows the reanalysed data for the mutant target assays, using a higher cut off of Ct36.5. Using this higher cut off resulted in 20/ 31 (64.5%) signals being designated as positive, an increase of 58.0% over the previous analysis.

Based on these preliminary observations, the decision was made to continue with these adjusted cut off values into the next phase of analysis.

Table 3.3.1.4 TaqMan Array control assay performance with a selection of samples at varying DNA concentrations. Ranked and colour scaled in order of gDNA input, lowest to highest. Ct values considered positive are below 36.5, highlighted in green/ blue. Assays failing for generate a Ct are represented as 'Undetermined' (highlighted red/ pink).

Sample No	DNA input (ng, total)	Target Name	Ct Mean
1	18.0	BRAF_475_mu	37.00
1	18.0	BRAF_476_mu	38.23
1	18.0	NRAS_564_mu	38.10
9	30.0	BRAF_477_mu	33.99
32	31.3	NRAS_573_mu	39.86
3	47.8	NRAS_584_mu	34.24
8	60.0	BRAF_476_mu	36.18
31	62.5	BRAF_476_mu	39.48
31	62.5	EGFR_6240_mu	39.32
28	75.0	KRAS_19404_mu	35.31
36	100.0	KRAS_553_mu	31.44
4	100.2	KRAS_553_mu	32.87
27	150.0	KRAS_19404_mu	34.08
35	239.0	NRAS_584_mu	33.42
7	250.0	BRAF_476_mu	35.53
10	250.0	KRAS_520_mu	34.62
12	250.0	NRAS_580_mu	31.93
14	250.0	EGFR_6240_mu	39.17
16	250.0	BRAF_474_mu	32.58
16	250.0	NRAS_564_mu	32.56
19	250.0	BRAF_476_mu	34.66
20	250.0	EGFR_6253_mu	36.26
20	250.0	NRAS_565_mu	37.77
21	250.0	BRAF_476_mu	36.25
24	250.0	BRAF_476_mu	37.32
33	300.0	BRAF_476_mu	36.38
2	302.8	NRAS_583_mu	38.95
26	450.0	KRAS_19404_mu	32.99
6	500.0	BRAF_476_mu	34.02
11	500.0	EGFR_6240_mu	36.99
11	500.0	KRAS_520_mu	32.99

3.3.2 Optimisation of sample loading

The second priority of the optimisation process was to establish the optimal DNA loading conditions for analysing FFPE DNA samples using the TaqMan Array. The previous stage (described in 3.2.1) gave some indication of the relative performance of the TaqMan Array at varying DNA concentrations, however this next phase would aim to determine the optimal DNA input amount.

28 previously analysed samples were obtained from the UHCW Pathology Tissue Bank, DNA extracted using the Promega Maxwell, quantified by Qubit then analysed the TaqMan Array assay at varying concentrations, ranging from 250 ng to 1000 ng. Manufacturer's guidance suggested a DNA input range of 10- 1000 ng gDNA, however for diagnostic use I believed determining an optimal DNA input was an essential step. When working with clinical samples, material is often limited, meaning it may not always be possible to input a high amount of DNA.

The data from these samples is shown in table 3.3.2.1 and figure 3.3.2.1. It is clear from the control and target Ct values and the amplification plots that increasing gDNA input beyond a certain threshold has an inhibitory effect on assay performance.

Table 3.3.2.1 Data from increasing DNA input. Two full cards were run per DNA concentration at 250 ng and 500 ng. The data is arranged to correspond to the card layout. Only one card was run with 1000 ng DNA loading. Data from the 1000 ng run is could not be analysed by the software due to run failure.

		250 ng				500 ng				1000 ng	
		Run 1		Run 2		Run 1		Run 2		Run 1	
Plate Channel	Control	Ct Mean	PASS/ FAIL	Ct Mean	PASS/ FAIL	Ct Mean	PASS/ FAIL	Ct Mean	PASS/ FAIL	Ct Mean	PASS/ FAIL
1	EGFR_rf	Undetermined	FAIL	38.500	FAIL	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL
	BRAF_rf	Undetermined	FAIL	37.175	FAIL	37.361	FAIL	Undetermined	FAIL	Undetermined	FAIL
	KRAS_rf	Undetermined	FAIL	39.575	FAIL	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL
	NRAS_rf	Undetermined	FAIL	33.230	PASS	34.321	PASS	35.952	FAIL	Undetermined	FAIL
2	EGFR_rf	34.683	PASS	37.076	FAIL	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL
	BRAF_rf	32.684	PASS	34.991	PASS	36.938	FAIL	39.858	FAIL	Undetermined	FAIL
	KRAS_rf	36.432	FAIL	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL
	NRAS_rf	31.645	PASS	31.589	PASS	32.212	PASS	34.314	PASS	Undetermined	FAIL
3	EGFR_rf	35.269	FAIL	Undetermined	FAIL	35.295	FAIL	Undetermined	FAIL	Undetermined	FAIL
	BRAF_rf	33.550	PASS	Undetermined	FAIL	33.476	PASS	Undetermined	FAIL	Undetermined	FAIL
	KRAS_rf	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL
	NRAS_rf	31.107	PASS	35.497	FAIL	30.543	PASS	36.929	FAIL	Undetermined	FAIL
4	EGFR_rf	33.598	PASS	39.451	FAIL	Undetermined	FAIL	36.527	FAIL	Undetermined	FAIL
	BRAF_rf	31.591	PASS	37.760	FAIL	Undetermined	FAIL	34.234	PASS	Undetermined	FAIL
	KRAS_rf	33.081	PASS	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL
	NRAS_rf	30.344	PASS	33.174	PASS	Undetermined	FAIL	30.739	PASS	Undetermined	FAIL
5	EGFR_rf	34.915	PASS	35.017	FAIL	36.655	FAIL	Undetermined	FAIL	Undetermined	FAIL
	BRAF_rf	33.787	PASS	33.147	PASS	34.851	PASS	Undetermined	FAIL	Undetermined	FAIL
	KRAS_rf	36.937	FAIL	37.591	FAIL	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL
	NRAS_rf	30.647	PASS	31.797	PASS	31.294	PASS	Undetermined	FAIL	Undetermined	FAIL
6	EGFR_rf	36.802	FAIL	35.113	FAIL	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL
	BRAF_rf	33.844	PASS	32.694	PASS	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL
	KRAS_rf	Undetermined	FAIL	37.659	FAIL	Undetermined	FAIL	Undetermined	FAIL	Undetermined	FAIL
	NRAS_rf	32.273	PASS	31.713	PASS	38.181	FAIL	Undetermined	FAIL	Undetermined	FAIL
7	EGFR_rf	35.335	FAIL	36.035	FAIL	35.688	FAIL	33.949	PASS	Undetermined	FAIL
	BRAF_rf	32.741	PASS	33.770	PASS	33.289	PASS	31.344	PASS	Undetermined	FAIL
	KRAS_rf	38.311	FAIL	Undetermined	FAIL	Undetermined	FAIL	37.932	FAIL	Undetermined	FAIL
	NRAS_rf	30.849	PASS	31.664	PASS	31.735	PASS	29.616	PASS	Undetermined	FAIL
	Mean Ct	**33.656**		**35.192**		**34.417**		**34.672**		**N/A**	
	Controls pass		**16**		**10**		**8**		**6**		**0**
	Pass rate				**46.43%**				**25.00%**		**0%**

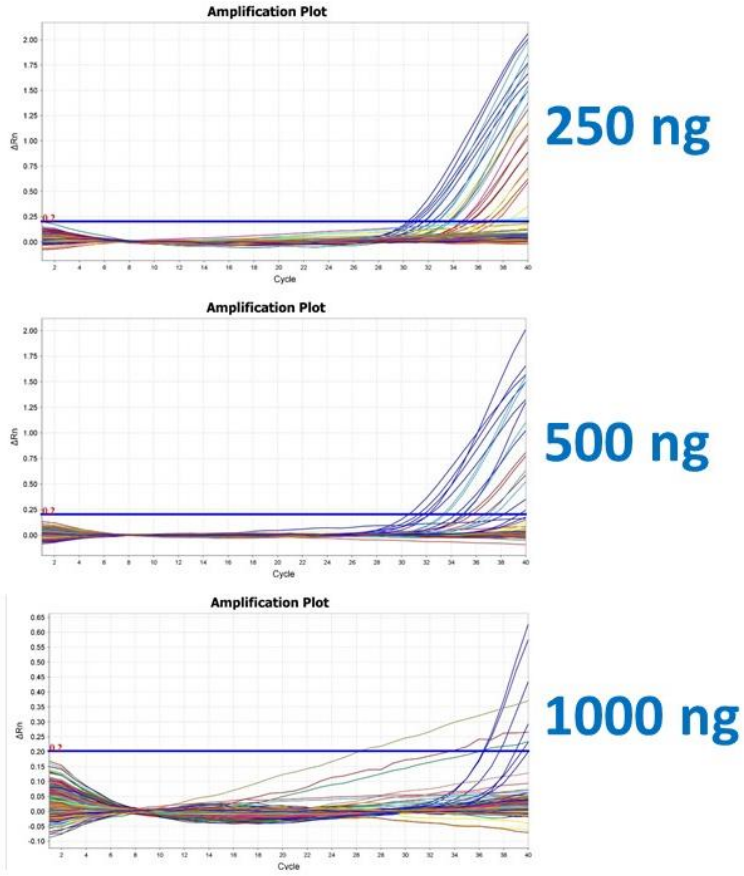

Figure 3.3.2.1 Amplification plots from TaqMan Array runs at varying gDNA concentrations. Increasing gDNA input seems to inhibit assay performance.

As a result of the data above, the next step was to perform serial dilutions at lower concentrations. Seven samples were serially diluted and analysed on the TaqMan Array. The samples were run at concentrations of 150, 100 and 50 ng of gDNA input. If assay performance was acceptable at these DNA inputs, then this would be very advantageous in the diagnostic setting due to the frequently low yield of DNA from certain specimens. The data from these serial dilutions is shown in figures 3.3.2.2- 5.

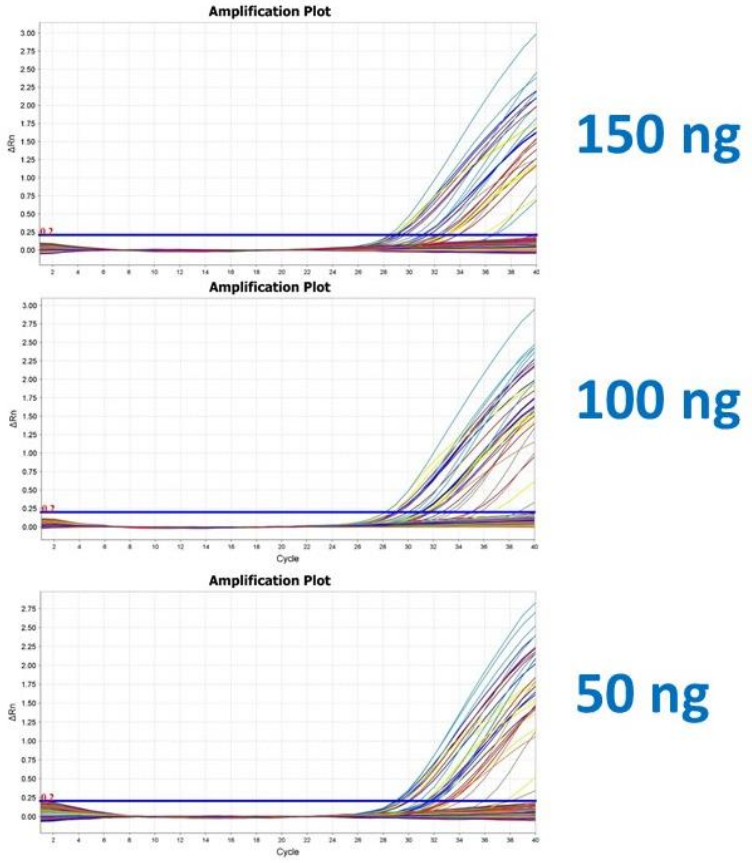

Figure 3.3.2.2 Amplification plots from TaqMan Array runs at varying gDNA concentrations. Decreasing gDNA input below 150 ng does not appear to inhibit assay performance.

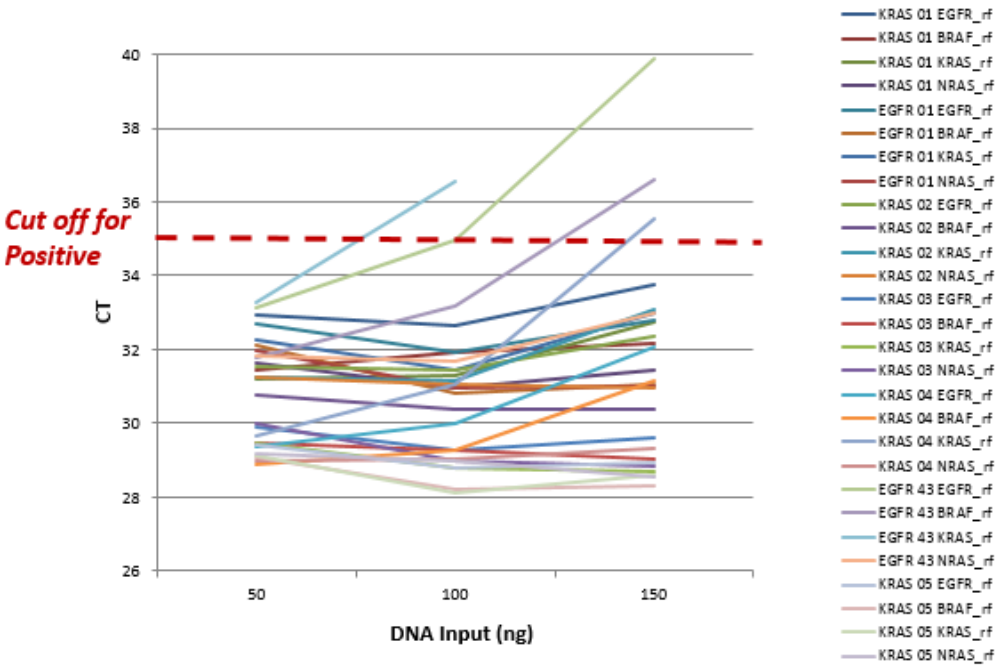

Figure 3.3.2.3 Performance of TA control assays (_rf) at varying DNA inputs. Series name have been removed

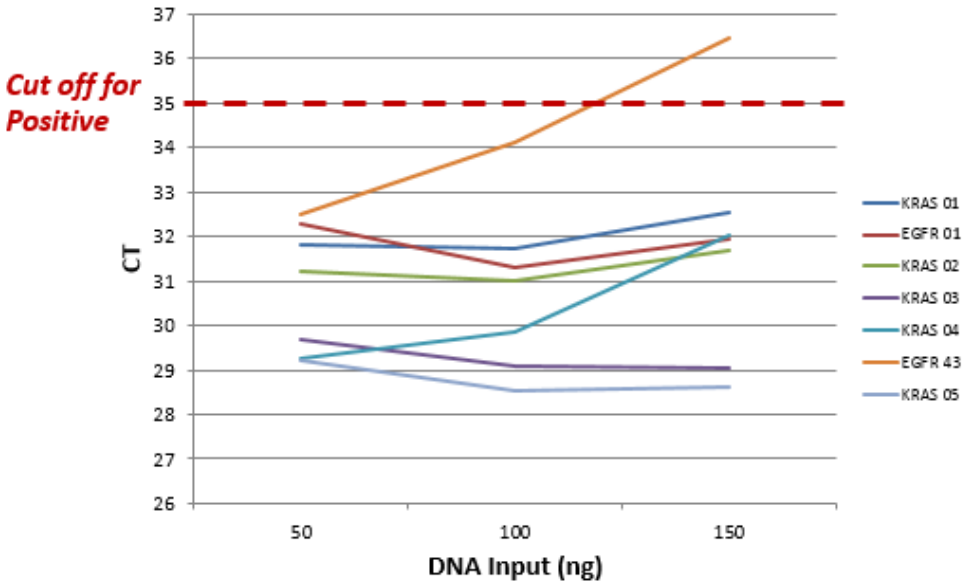

Figure 3.3.2.4 Summary chart of mean Ct from all control assays (_rf) at varying DNA inputs

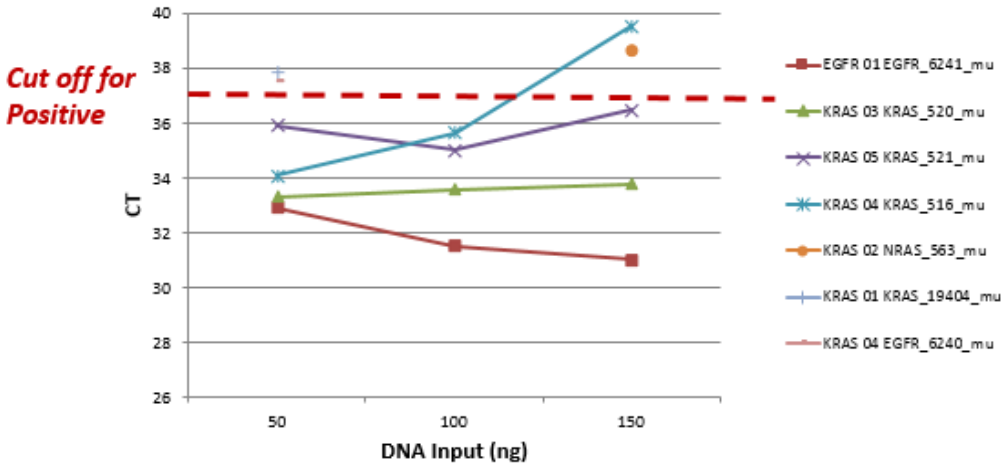

Figure 3.3.2.5 Performance of TA mutant target assays (_mu) at varying DNA inputs.

Figures 3.3.2.2- 5 show the performance of the 7 serially diluted samples from 50- 150 ng. The increase in Ct as DNA input increases can be easily seen in the amplification plots shown

in figure 3.3.2.2. This suggests a slight inhibitory effect with increased DNA loading. Figure 3.3.2.3 shows the Ct values of the control assays in more detail. Whilst many signal remain stable from 50- 150 ng of DNA input, there are a number of samples where Ct increases, and one example where the assay failed at 150 ng DNA input. The average Ct values for all the control assays generated by each sample was calculated and plotted in figure 3.3.2.4. This figure clearly shows the trend of increasing Ct with increased DNA input.

With the mutant target assays the relationship between DNA input and assay performance was not as clear. The data is shown in figure 3.3.2.5. In some cases increasing DNA input lower Ct (sample EGFR 01) which reflects improved assay performance. However in other cases (KRAS 03, KRAS 04 and KRAS 05) increasing DNA input resulted in higher Ct, suggesting reduced assay performance.

Due to the limitations in terms of the number of samples available, and the often low DNA concentrations experienced with clinical samples, the final analysis parameters were decided based on the observations described in sections 3.3.1 and 3.3.2.

Final Analysis Thresholds:

Sample loading: 50 ng total gDNA input.

Control assays: Ct<=35 for positive.

Mutant Target assays: Ct<=36.5 for positive.

3.3.3 Parallel testing of Pathology Department Diagnostic Samples

Once optimal DNA loading and analysis cut offs had been determined, the next phase of this validation was to retest a panel of previously analysed samples using the TaqMan Array.

3.3.4 Parallel testing: EGFR

42 FFPE tissue samples from NSCLC patients were obtained from UHCW Pathology, DNA was extracted from micro dissected sections and the DNA analysed using the TaqMan Array. Table 3.2.4.1 shows the data from the NSCLC samples, including the original Therascreen results and the new results generated by the TaqMan Array.

Table 3.3.4.1 Results of NSCLC validation samples, including Therascreen and TaqMan Array results. WT = Wild-type.

Sample ID	PCR PASS/ FAIL	Positive EGFR Targets	Ct	TA Result	Therascreen Result	Concordance
EGFR 01	PASS	EGFR_6241_mu	32.905	S768I	S768I	MATCH
EGFR 02	PASS	EGFR_ex19dels_mu	28.986	Exon 19 Del	Exon 19 Del	MATCH
EGFR 03	PASS	EGFR_6224_mu	31.630	L858R	L858R	MATCH
EGFR 04	PASS	EGFR_ex19dels_mu	36.119	Exon 19 Del	Exon 19 Del	MATCH
EGFR 05	PASS	None		WT	WT	MATCH
EGFR 06	PASS	None		WT	WT	MATCH
EGFR 07	PASS	None		WT	WT	MATCH
EGFR 08	PASS	EGFR_ex19dels_mu	32.152	Exon 19 Del	Exon 19 Del	MATCH
EGFR 09	PASS	EGFR_ex19dels_mu	30.311	Exon 19 Del	Exon 19 Del	MATCH
EGFR 10	PASS	EGFR_6213_mu; EGFR_6252_mu	33.492; 33.841	L861Q and G719X	L861Q and G719X	MATCH
EGFR 11	PASS	EGFR_ex19dels_mu	31.994	Exon 19 Del	Exon 19 Del	MATCH
EGFR 12	PASS	None		WT	WT	MATCH
EGFR 13	PASS	EGFR_6213_mu	31.691	L861Q	L861Q	MATCH
EGFR 14	PASS	EGFR_ex19dels_mu	30.878	Exon 19 Del	Exon 19 Del	MATCH
EGFR 15	PASS	EGFR_ex19dels_mu	31.060	Exon 19 Del	Exon 19 Del	MATCH
EGFR 16	PASS			WT	WT	MATCH
EGFR 17	PASS			WT	WT	MATCH
EGFR 18	PASS	EGFR_ex19dels_mu	30.501	Exon 19 Del	Exon 19 Del	MATCH
EGFR 19	PASS	None		WT	WT	MATCH
EGFR 20	PASS	None		WT	WT	MATCH
EGFR 24	PASS	None		WT	WT	MATCH
EGFR 25	PASS	None		WT	WT	MATCH
EGFR 26	PASS	None		WT	WT	MATCH
EGFR 27	PASS	None		WT	WT	MATCH
EGFR 28	PASS	None		WT	WT	MATCH
EGFR 29	PASS	None		WT	WT	MATCH
EGFR 30	PASS	None		WT	WT	MATCH
EGFR 31	PASS	None		WT	WT	MATCH
EGFR 32	PASS	EGFR_ex19dels_mu	34.947	Exon 19 Del	Exon 19 Del	MATCH
EGFR 33	PASS	None		WT	WT	MATCH
EGFR 34	PASS	None		WT	WT	MATCH
EGFR 35	PASS	None		WT	WT	MATCH
EGFR 36	PASS	None		WT	WT	MATCH
EGFR 37	PASS	None		WT	WT	MATCH
EGFR 38	PASS	None		WT	WT	MATCH
EGFR 39	PASS	None		WT	WT	MATCH
EGFR 40	PASS	None		WT	WT	MATCH
EGFR 41	PASS	None		WT	WT	MATCH
EGFR 42	PASS	None		WT	WT	MATCH
EGFR 21	PASS	None		WT	S768I	DISCREPANT
EGFR 22	PASS	None		WT	Exon 20 Ins	DISCREPANT
EGFR 23	PASS	None		WT	T790M	DISCREPANT

39/42 samples matched the original Therascreen results, demonstrating a high level of concordance (92.86%). All samples gave strong Ct values for the control assays (<35), indicating well optimised control robust reactions with good tolerance for FFPE DNA which can be variable in quality from sample to sample.

There were 3 discrepancies: all were detected as wild-type by TaqMan Array but were previously positive on Therascreen. To confirm the mutation status of these samples they were analysed using the Ion Torrent NGS platform. The results are given in table 3.3.4.2. In two cases (EGFR 21 and 23) the NGS analysis confirmed the TaqMan results, so these were not true discrepancies, and the Therascreen results were incorrect. In the final discrepant sample (EGFR 22) NGS analysis confirmed the presence of an exon 21 insertion, however this was not a true discrepancy as the panel of mutations included in our TaqMan Array plate design did not include this exact mutation, therefore I would not expect the TaqMan Array to be able to detect it. NGS Analytics show the exon 20 insertion to have been at high frequency (56.7%) but relatively low coverage (216 reads) so this mutation may not have constituted a large proportion of the original sample. This discrepancy occurred due to differences in design between the two assays, and not due to an issue to sensitivity or specificity. So after confirmation by NGS, the adjusted concordance between Therascreen and TaqMan Array was 41/42 (97.62%).

Table 3.3.4.2 NSCLC samples giving discrepant results (Therascreen vs TaqMan Array) and associated NGS results. WT = Wild-type.

Sample ID	PCR Results			NGS Data			Designation
	TA	Therascreen	Concordance	NGS Result	Coverage	Frequency	
EGFR 21	WT	S768I	DISCREPANT	WT	-	-	True Negative
EGFR 22	WT	Exon 20 Ins	DISCREPANT	Exon 20 Ins	216	56.7	False Negative
EGFR 23	WT	T790M	DISCREPANT	WT	-	-	True Negative

3.3.5 Parallel testing: KRAS/ NRAS

26 FFPE tissue samples from NSCLC patients were obtained from UHCW Pathology, DNA was extracted from macrodissected sections and the DNA analysed using the TaqMan Array.

Table 3.3.5.1 shows the samples where the original castPCR results matched the new results generated by the TaqMan Array. 7 samples had to be excluded from the analysis due to poor sample quality, therefore only 19 were included in the *KRAS*/ *NRAS* validation.

Table 3.3.5.1 Results of colorectal validation samples, including castPCR and TaqMan Array results. WT = Wild-type.

Sample ID	PCR PASS/ FAIL	Positive Targets	Ct	TA Result	CAST Result	Concordance
KRAS 01	PASS	None		WT	WT	MATCH
KRAS 02	PASS	None		WT	WT	MATCH
KRAS 03	PASS	KRAS_520_mu	33.321	G12V	G12V	MATCH
KRAS 04	PASS	KRAS_516_mu	34.060	G12C	G12C	MATCH
KRAS 05	PASS	KRAS_521_mu	35.893	G12D	G12D	MATCH
KRAS 06	PASS	KRAS_532_mu	31.412	G12D	G12D	MATCH
KRAS 07	PASS	KRAS_532_mu	35.956	G12D	G12D	MATCH
KRAS 09	PASS	NRAS_580_mu	31.751	Q61L	Q61L	MATCH
KRAS 10	PASS	KRAS_522_mu	34.899	G12A	G12A	MATCH
KRAS 11	PASS	KRAS_521_mu	33.179	G12D	G12D	MATCH
KRAS 12	PASS	NRAS_564_mu	33.023	G12D	G12D	MATCH
KRAS 13	PASS	KRAS_522_mu	35.350	G12A	G12A	MATCH
KRAS 14	PASS	KRAS_521_mu	34.220	G12D	G12D	MATCH
KRAS 15	PASS	KRAS_521_mu	35.144	G12D	G12D	MATCH
KRAS 16	PASS	None		WT	WT	MATCH
KRAS 17	PASS	None		WT	WT	MATCH
KRAS 18	PASS	None		WT	WT	MATCH
KRAS 19	PASS	None		WT	WT	MATCH
KRAS 08	PASS	None		WT	G12D	DISCREPANT

The TA assay shows a high degree of concordance with CastPCR results, 18/19 (94.74%). There was one discrepant sample (KRAS 08) which was determined to be wild-type by TA but was determined to contain a G12D mutation by the original castPCR assay. This sample was sequenced using the Ion Torrent NGS system and the G12D mutation was detected. Therefore this was a true discrepancy, where the mutation was not detected by TA but successfully detected by castPCR and NGS.

Table 3.3.5.2 Colorectal samples giving discrepant results (castPCR vs TaqMan Array) and associated NGS results. WT = Wild-type.

Sample ID	TA Result	CAST Result	TA vs CAST	NGS Result	Designation
KRAS 08	WT	G13D	DISCREPANT	G14D Detected	False Negative

3.3.6 Parallel testing: *BRAF*

28 FFPE tissue samples from melanoma patients were obtained from UHCW Pathology, DNA was extracted from micro dissected sections and the DNA analysed using the TaqMan Array and Ion Torrent NGS. Table 3.3.6 shows the results of these tests. The TaqMan Array results from 27 of the 28 samples matched the results generated by NGS. There was one discrepant sample (MLM 27) where the TaqMan Array detected a V600E mutant which was not detected by NGS. This was determined to be a true discrepancy.

Table 3.3.6.1 Results of Melanoma validation samples, including TaqMan Array and NGS results. WT = Wild-type.

Study ID	PCR PASS/ FAIL	Positive Targets	Ct	TA RESULT	NGS	Concordance
MLM 01	PASS	NRAS_580_mu	31.714	Q61K	Q61K	MATCH
MLM 02	PASS	BRAF_476_mu	32.078	V600E	V600E	MATCH
MLM 03	PASS			WT	WT	MATCH
MLM 04	PASS	BRAF_476_mu	35.811	V600E	V600E	MATCH
MLM 05	PASS	BRAF_476_mu	33.455	V600E	V600E	MATCH
MLM 06	PASS	NRAS_584_mu	34.406	Q61R	Q61R	MATCH
MLM 07	PASS	NRAS_584_mu	34.056	Q61R	Q61R	MATCH
MLM 08	PASS			WT	WT	MATCH
MLM 09	PASS	NRAS_584_mu	35.556	Q61R	Q61R	MATCH
MLM 10	PASS			WT	WT	MATCH
MLM 11	PASS			WT	WT	MATCH
MLM 12	PASS			WT	WT	MATCH
MLM 13	PASS	NRAS_584_mu	35.592	Q61R	Q61R	MATCH
MLM 14	PASS	BRAF_476_mu	32.146	V600E	V600E	MATCH
MLM 15	PASS	NRAS_584_mu	36.232	Q61R	Q61R	MATCH
MLM 16	PASS	BRAF_476_mu	33.090	V600E	V600E	MATCH
MLM 17	PASS	BRAF_476_mu	31.873	V600E	V600E	MATCH
MLM 18	PASS			WT	WT	MATCH
MLM 19	PASS	BRAF_476_mu	33.260	V600E	V600E	MATCH
MLM 20	PASS			WT	WT	MATCH
MLM 21	PASS	BRAF_476_mu	36.159	V600E	V600E	MATCH
MLM 22	PASS	BRAF_476_mu	32.099	V600E	V600E	MATCH
MLM 23	PASS	BRAF_476_mu	36.474	V600E	V600E	MATCH
MLM 24	PASS			WT	WT	MATCH
MLM 25	PASS	BRAF_476_mu	34.339	V600E	V600E	MATCH
MLM 26	PASS	NRAS_569_mu	35.205	G13R	G13R	MATCH
MLM 28	PASS	NRAS_584_mu	35.093	Q61R	Q61R	MATCH
MLM 27	PASS	BRAF_475_mu	31.447	V600E	WT	DISCREPANT

Table 3.3.6.2 Colorectal samples giving discrepant results (castPCR vs TaqMan Array) and associated NGS results. WT = Wild-type.

Study ID	TA Result	NGS Result	TA vs NGS	Designation
MLM 27	V600E	WT	DISCREPANT	False Positive

3.3.7 Sensitivity and Specificity

The TaqMan array assay performed well with a majority of the lung, colorectal and melanoma samples used for validation. A summary of the data and associated statistics for sample set are shown in tables 3.3.7.1- 3. A summary of the overall data is shown in table 3.3.7.4. Overall the assay demonstrated a sensitivity of 95.7% and a specificity of 97.8%.

Table 3.3.7.1 Validation summary of lung samples tested using the TaqMan Array assay.

TA Validation: Lung				
Test	**Mutant Present**		**Mutant absent**	
Positive	True Positive	13	False Positive	0
Negative	False Negative	1	True Negative	28

Statistic	**Value**
Sensitivity	**92.9%**
Specificity	**100.0%**
Positive Predictive Value	**100.0%**
Negative Predictive Value	**96.6%**

Table 3.3.7.2 Validation summary of colorectal samples tested using the TaqMan Array assay.

TA Validation: Colorectal				
Test	**Mutant Present**		**Mutant absent**	
Positive	True Positive	12	False Positive	0
Negative	False Negative	1	True Negative	6

Statistic	**Value**
Sensitivity	**92.3%**
Specificity	**100.0%**
Positive Predictive Value	**100.0%**
Negative Predictive Value	**85.7%**

Table 3.3.7.3 Validation summary of melanoma samples tested using the TaqMan Array assay.

TA Validation: Melanoma				
Test	**Mutant Present**		**Mutant absent**	
Positive	True Positive	19	False Positive	1
Negative	False Negative	0	True Negative	10

Statistic	**Value**
Sensitivity	**100.0%**
Specificity	**90.9%**
Positive Predictive Value	**95.0%**
Negative Predictive Value	**100.0%**

Table 3.3.7.4 Validation summary of Melanoma samples tested using the TaqMan Array assay.

TA Validation: All Sample types				
Test	**Mutant Present**		**Mutant absent**	
Positive	True Positive	44	False Positive	1
Negative	False Negative	2	True Negative	44

Statistic	**Value**
Sensitivity	**95.7%**
Specificity	**97.8%**
Positive Predictive Value	**97.8%**
Negative Predictive Value	**95.7%**

3.3.8 Prospective testing of plasma cfDNA

After the validation of the TA using DNA extracted from lung, colorectal and melanoma tissue samples (sections 3.3.4, 3.3.5 and 3.3.6), the TA assay was implemented as the diagnostic assay for cancer gene mutations detection in *EGFR*, *KRAS*, *NRAS* and *BRAF* in the UHCW Pathology Department. Shortly after implementation I began prospectively testing plasma cfDNA samples submitted for testing by clinical staff. Plasma cfDNA samples were prepared as described in methods and loaded into the TA cards to the same specifications as tissue (up to 50 ng DNA total load per channel). To date 16 cfDNA samples have been tested using the TA. The results from these runs are shown in table 3.2.8.

Table 3.3.8 TA results for control and mutant assays using prospective cfDNA samples. Control assays giving a Ct below 35 are highlighted in blue/green. DNA loading concentrations below 50 ng are highlighted in red. WT = Wild-type.

STUDY ID	DNA loading (ng)	EGFR Ref Ct	EGFR	BRAF Ref Ct	BRAF	KRAS Ref Ct	KRAS	NRAS Ref Ct	NRAS
TA Plasma 001	50.0	27.74	WT	27.64	WT	27.44	WT	28.37	WT
TA Plasma 002	50.0	27.86	WT	27.48	WT	27.73	WT	28.28	WT
TA Plasma 003	11.2	30.58	WT	30.43	WT	30.78	WT	31.29	WT
TA Plasma 004	14.3	31.10	WT	30.15	WT	30.45	WT	30.70	WT
TA Plasma 005	50.0	27.98	WT	27.94	WT	27.89	WT	28.04	WT
TA Plasma 006	13.3	30.06	WT	29.66	WT	29.93	WT	29.98	WT
TA Plasma 007	9.1	30.97	WT	30.29	WT	31.11	WT	31.24	WT
TA Plasma 008	26.6	30.04	WT	29.47	WT	29.52	WT	30.27	WT
TA Plasma 009	<50.0	32.37	WT	31.27	WT	32.36	WT	32.57	WT
TA Plasma 010	44.7	27.70	WT	28.11	WT	27.97	WT	28.20	WT
TA Plasma 011	50.0	28.02	WT	27.66	WT	27.68	WT	28.03	WT
TA Plasma 012	13.2	30.02	WT	29.56	WT	30.59	WT	30.13	WT
TA Plasma 013	50.0	27.66	E709K Ct=37.907	27.28	WT	27.65	WT	27.91	WT
TA Plasma 014	7.5	30.90	WT	30.97	WT	32.11	WT	30.84	G13S; c.37G>A Ct=36.884
TA Plasma 015	5.4	31.48	WT	31.05	WT	31.88	WT	31.53	WT
TA Plasma 016	<50.0	31.95	WT	31.91	WT	31.83	WT	32.46	WT
Median		**30.05**		**29.61**		**30.19**		**30.20**	
Min		**27.66**		**27.28**		**27.44**		**27.91**	
Max		**32.37**		**31.91**		**32.36**		**32.57**	

Due to the characteristics of cfDNA only five of 16 samples could be loaded at 50 ng per channel. The samples varied greatly in their DNA concentrations and how much DNA could subsequently be loaded into the TA card. Two samples (TA Plasma 009 and 016) were at such low concentration that they could not be quantified by the Qubit instrument, so as a result the exact amount of DNA loaded for those samples is not known.

Despite these low DNA concentrations, the cfDNA samples all gave strong positive signals for all the control assays, there were no failures (Ct>35). The control assay Ct ranged from 27.28 to 32.57, all a comfortable margin below the cut off of 35. No mutant assays gave a signal

below our current cut off (Ct<36.5), however two mutant assays (*EGFR* E709K and *NRAS* G13S) gave Cts with two samples (TA Plasma 013 and 014). Unfortunately the Ct values were above our threshold for positive samples (Ct37.907 and Ct36.884) but it suggests there is potential for cfDNA to be a suitable analyte for use with the TA format of assay.

3.4 DISCUSSION

In general there was very good concordance between the TA, Therascreen, castPCR and NGS. 88 of the 91 (96.7%) samples tested agreed with the original results. There were only three true discrepancies (one false negative in *EGFR* and *BRAF*, one false positive in *KRAS*). Whilst it is unfortunate that these discrepancies occurred, some discrepancies would be expected when compared different techniques due to the individual characteristics and analysis parameters of the various assays. The presence of two false negatives may suggest that the TA assay has slightly lower sensitivity than some of the other assays, however the impact of this does not seem to be significant.

NGS proved to be an invaluable tool for this validation process, giving very accurate and sensitive results for confirming the mutation status of discrepant samples. However it is expensive and time consuming, and therefore a PCR based method offers good sensitivity and specificity as well as being rapid and cost effective. The minor compromise in this case is some loss of sensitivity or specificity compared to NGS, however these differences can be partially compensated for by quality control measures at the DNA extraction and quantification phase i.e. maximising sample quality prior to analysis on the TA.

An advantage of the TA is the very low DNA input requirement. During the optimisation phase it was found that the TA control and mutant target assays worked consistently well with only 50 ng of DNA. Given the clinical context, in particular with EGFR samples which are often extracted from FFPE slides, then having a robust assay with very low DNA input requirements is a considerable advantage. The Therascreen assay manufacturer guidelines do not specify an exact DNA input requirement, however the assay does require eight separate reactions to be run which uses a large volume of the extracted DNA sample. The

castPCR assay for *KRAS* requires 50 ng DNA per reaction, and therefore 800 ng DNA in total. The low DNA requirement of TA assay is a major advantage as it not only means dilute samples can be analysed (which otherwise might not be possible with the other assays) but also repeat analysis may also be possible of there is remaining material. Again with reference to the clinic, this means the TA may allow results to be obtained for samples which would otherwise fail, and thus the genetic information would not be available for the management of the patient.

The TA assay had another major advantage over the other assays in terms of the number of pipetting steps and the hands on time. After sample preparation (diluting sample DNA with water and TaqMan PCR Master Mix), the entire card can be loaded with just one pipetting action per channel (therefore a maximum of eight actions). This gives a great advantage in terms of staff time but also reduces the chances of accidently contamination by operator error. Subsequently this will have a positive effect of the failure rate of runs, and thus improvements in service turnaround time and reagent costs.

Following this validation study, the Pathology Department at UHCW has implemented the TA assay (designed and validated here) as their diagnostic test for *EGFR*, *KRAS*, *NRAS* and *BRAF* mutation detection of NSCLC, CRC and melanoma samples. It has been a great success, readily adopted by the Molecular Laboratory, and is routinely run approximately twice a week. The combining of two assays into one has generated a saving in staff time.

Alongside the routine tissue testing using the TA assay, there has also been prospective pilot testing of plasma cfDNA using the TA. The results from these tests have shown that although no positive mutants were detected (within our analysis parameters) the control assays performed well within a consistent range, which strongly suggests that cfDNA has potential

as a suitable source of DNA for testing in the TA. It is particularly impressive that the control assays gave strong results even with very low DNA inputs (<10 ng in a number of cases). The weakness of this approach is that the tumour DNA (containing the target mutations) is at very low concentrations and very highly diluted in wild-type genomic DNA, so the assay in its current format is not able to detect those mutations consistently. The low DNA concentrations in the samples also meant it was not possible to confirm the mutation status of the samples by NGS.

A frequent challenge with cfDNA is the low concentration of DNA in the plasma, which then has an impact on the amount of DNA which can be loaded into the TA card. It may be possible to partially overcome this technical challenge by extracting a large volume of EDTA blood, but this is dependent on clinical and logistical factors associated with the blood collection. In terms of future validation, assuming there is a good source of plasma samples, it may be beneficial to optimise the DNA loading of cfDNA samples on TA. Increasing the DNA concentration may increases the abundance of the tumour DNA in the sample, potentially allowing them to be detected by the TA chemistry. However increasing the DNA loading of cfDNA samples may have a similar inhibitor effect on the PCR (as seen this FFPE DNA in this chapter), but this can only be established experimentally.

A practical consideration for diagnostic assays is cost. In order to introduce the TA assay into routine use in the UHCW Pathology Department a business case was submitted to the Trust, which included a financial assessment of the projected cost of running the assay based on predicted numbers of samples. According to this analysis, assuming 450 tests per year, the TA assay was projected to cost £183 per test, compared to £233 per test for Therascreen, a saving of £50 per test. A summary of these costings can be found in the chapter 3 appendix.

This chapter has shown the TA assay to be a robust and reliable approach for the detection of cancer gene mutations from extracted FFPE tissue DNA. It is also rapid and user friendly to set up, and its relatively broad range of targets (by PCR standards) makes it an efficient and cost effective solution for the service demands of a molecular diagnostic laboratory.

There are a number of ways this technology can be developed in the future. Firstly, the molecular laboratory (UHCW) continues to run a number of molecular assays on a number of different platforms (Qiagen Rotorgene®, Life Technologies ViiA7®, Abbott M2000 RT®, Cepheid SmartCycler® and Roche LightCycler®). Running and maintaining a suite of different PCR instruments is costly and requires a large amount of bench space, and also requires the staff to be familiar with numerous assays that are essentially very similar. If custom TaqMan Array plates could be designed and validated it would greatly increase efficiency within the molecular laboratory, and reduce the bench space required to house multiple PCR instruments.

Secondly, more specific to cancer mutation detection, the TA technology could be utilised to design more custom plates for detecting targets in other cancers e.g. ovarian, pancreatic, breast etc. This would greatly increase the services the Pathology department could offer without a significant logistical impact as no new equipment would be required, only a supply for the TA cards.

In terms of the plasma testing, this remains a significant challenge for molecular diagnostics. One approach could be to experiment with higher DNA concentrations (described earlier). Another could be the development of a pre-amplification step of the extracted DNA before loading into the TA i.e. a multiplex PCR to amplify the target genes. This would of course add

additional time and cost to the assay, however it would most likely still be cheaper than using an NGS approach.

The limitations of this work are associated with the format of the technology that is being utilising and the logistics of working in a pathology laboratory. The TA, whilst robust and reliable, can only detect 44 targets per channel, and whilst this is an improvement on many other assays, it inevitably means not all targets can be covered, although the most common are included in the panel. In terms of the samples used, ideally more samples would have been used in the validation stage, although 91 was sufficient. This is not an issue that could be controlled by our research group. Finally, more plasma samples and a large volume of each sample would have been very beneficial to the prospective plasma testing phase, although this again is an issue that was controlled by external factors.

A3 CHAPTER THREE APPENDIX

Table A3: Costings estimate for new TaqMan Array assay (referred to here as 'REB Array') compared to the currently in use test (Qiagen Therascreen). Adapted from UHCW Pathology Department TaqMan Array business case.

	Summary	Unit cost	Number	Total
TaqMan Array	**REB Array (New test)**			
	Cost of consumables per test	54.14	450	24,363
	External quality assurance	3.00	836	2,509
	Equipment rental and maintenance	1.00	15,710	15,710
	Training	*49.48*	5	247
	Staff costs	1.00	18,151	18,151
	Overheads @ 35%		1	21,343
	Total			**82,324**
	Cost per patient			**183**
Therascreen	**Current Test (450 tests)**			
	Cost of consumables per test	150.00	450	67,500
	External quality assurance	3.00	1,931	5,793
	Equipment rental and maintenance	1.00	3,388	3,388
	Training	150.00	5	750
	Staff costs	1.00	27,260	27,260
	Overheads @ 35%		1	36,642
	Total			**104,691**
	Cost per patient*			**233**
	*Only one gene per cancer - in many cases 2-3 will be needed			

Chapter Four – Automated isolation of circulating cell free DNA from blood plasma in lung cancer

4.1 INTRODUCTION

Circulating Free DNA (cfDNA) is described as non-cell associated DNA found in various bodily fluids (lymph, plasma, semen, saliva). It is generated by natural and disease associated cellular degradation; as cells are lysed by apoptosis, autophagy, or necrosis, their nucleic acids will be released from the nucleus and cytoplasm to enter the surrounding environment (Qin *et al.*, 2016). Exosomes are also known to be a source of cfDNA (Hyun *et al.*, 2016). Upon entering the surrounding environment, the blood plasma for example, the cfDNA has a number of fates. Firstly it can be degraded by natural DNase enzymes circulating in the plasma. Secondly, it can be removed from the blood stream by the liver, spleen or kidneys (Volik *et al.*, 2016). Elimination of cfDNA is part of host natural immune system and as a result elimination of cfDNA is efficiently performed, resulting in the half-life of cfDNA in plasma being very short, as low as 16 min reported in one study (Lo *et al.*, 1999). Other studies have found the clearance kinetics of cfDNA to be slightly longer at around 13 hours (Yu *et al.*, 2013), however it is certain that cfDNA turnover *in vivo* is a relatively rapid process. More research in this area is needed to ensure that cfDNA assays for diagnostic use are clinically reliable.

cfDNA was first identified in the 1940s (Mandel & Metais, 1948) and also described in the 1970s (Leon *et al.*, 1977), however it is only recently that interest has grown in utilising cfDNA as a potential diagnostic tool (Johnson & Lo, 2002). A number of studies have described elevated cfDNA in disease sufferers compared to healthy controls (Mead *et al.*, 2011; Szpechcinski *et al.*, 2012; Szpechcinski *et al.*, 2015) so some researchers suggested that crude cfDNA levels alone can be used as diagnostic disease biomarker, however other studies have found cfDNA levels to be highly variable both between individuals and between malignant

and benign conditions. As a result cfDNA levels alone cannot be used as a confirmation of malignant disease (Shaw *et al.*, 2012) until our understanding is more complete.

A major challenge in lung cancer is to obtain tissue samples suitable for molecular analysis, as the anatomy of the lung often means tumours cannot be biopsied until a fairly advanced stage, at which point patient outcomes become more unfavourable and therapy becomes less effective. In this context, mutation detection using cfDNA potentially offers considerable advantage over tissue in terms of how early oncogenic mutations can be identified, and subsequently at what stage the patient begins treatment.

A number of recent studies have investigated the potential use of cfDNA for cancer mutation detection. A study by Shaw *et al.* (Shaw *et al.*, 2012) examined the use of cfDNA to detect gene copy number variation (CNV) in breast cancer patients using Affymetrix SNP (Single Nucleotide Polymorphism) arrays. The study showed that CNV in specific genes could be detected in cfDNA, and that genetic analysis of cfDNA could distinguish cancer patients from healthy controls. Certain disease markers were found to be present in the cfDNA even after successful treatment or surgery, which implies cfDNA could be potentially used as a means of monitoring disease dormancy.

There have been a number of studies specifically investigating the use of cfDNA for mutant detection in lung cancer patients. Typically concordance between tumour DNA and cfDNA has been found to be encouragingly high. The study by Ishii *et al.* (Ishii *et al.*, 2015) compared the detection of four types of lung cancer mutants in cfDNA and tissue using digital PCR. The findings were that the overall concordance between cfDNA and tissue was 83.3%, and that cfDNA had potential as a means of detecting genetic changes contributing to TKI resistance in patients receiving therapy. Another study by Mok *et al.* (Mok *et al.*, 2015) analysed cfDNA

as part of a phase III clinical trial of erlotinib with platinum based chemotherapy. Blood samples were collected before, during and after the treatment phases, and molecular analysis of the isolated cfDNA gave both an insight into the concordance between blood and tissue DNA, and gave an indication of the dynamics of the genetic changes occurring as a result of treatment. Multiple studies are consistently proposing a strong case for the use of cfDNA in cancer diagnostics and disease monitoring.

Biotechnology companies are starting to appreciate the potential clinical utility of cfDNA, and as a result a number of kits have become available for cfDNA isolation, either by adaptation of an existing kit or specific new kits have been designed and manufactured. Manual and automated solutions are available. A number of recent studies have evaluated manual methods (Sherwood *et al.*, 2016; Yuan *et al.*, 2012), however recent guidance advises that automated methodologies are far more suited to the routine diagnostic setting (Cree *et al.*, 2014). Another equipment issue related to cfDNA are the type of tubes used to collect the whole blood from the patient. A recent study investigated specifically designed blood collection tubes for the stabilization of cfDNA before plasma separation (Toro *et al.*, 2015). Toro *et al.* compared Cell-free DNA BCT tubes (Streck, NE) and PAXgene DNA blood tubes (Qiagen). They concluded the Streck tubes were effective at stabilizing cell free DNA in plasma for up to a week at room temperature. Whilst using specifically designed blood tubes would be ideal, introducing a new tube type into the UHCW phlebotomy department simply would not be practical. Performing this investigation in the diagnostic Pathology laboratory setting presents a number of limitations on sample types and equipment available.

I aimed to perform a side-by-side comparison of two leading platforms for automated cfDNA extraction from plasma: the commonly available Qiagen EZ1 instrument used frequently to isolate circulating viral DNA in plasma, running the DSP Virus Kit 2.0 (Qiagen), and the

Promega Maxwell instrument running the prototype Circulating DNA Kit (Promega). It is important to state clearly that the Qiagen DSP Virus Kit is not specifically designed for cfDNA isolation from plasma, however there was no suitable alternative prototype or retail kit available from Qiagen at the time of this study. The kit chemistry is designed to isolate viral DNA from plasma, so it should in theory be able to isolate cfDNA from plasma also. The Qiagen and Promega extraction platforms were selected for this study as they were routinely used for DNA extraction in the investigating laboratory.

4.1.1 Aims

There are three aims for this study:

- To determine if cfDNA can be isolated from standard EDTA blood samples, using existing equipment used in the diagnostic laboratory.

- To establish which of the available standard extraction platforms is most efficient at extracting cfDNA, and which yields the best quality DNA for downstream applications.

- To evaluate if cfDNA obtained from these samples and extraction platforms can be used to detect cancer mutations, using pre-designed NGS assays designed for use with tissue.

4.2 METHODS

4.2.1 Equipment

Bench top centrifuge compatible with 4ml EDTA Blood tubes (up to 1600g)

Bench top centrifuge compatible with 1.5ml micro-centrifuge tubes (up to 15500g)

Bench top centrifuge compatible with 96 well PCR plates (up to 1600g) and Taqman arrays

Bench top centrifuge compatible with 0.2 ml PCR strip tubes

Bench top centrifuge compatible with Ion Torrent sequencing chips (314, 316 and 318)

Heat block capable of heating 1.5ml tubes to 70°C

Qiagen EZ1 automated extraction instrument (see figure. 2.3)

Promega Maxwell Automated Extraction Instrument (see figure. 2.3)

Laminar Flow safety cabinet with UV decontamination feature

Nanodrop spectrophotometer (see figure 4.2.1.2)

Qubit 2.0 Nucleic Acid Analyser (see figure 2.2.7)

Life Technologies ViiA 7 Dx Real Time PCR instrument with Fast 96 well block

Life Technologies ViiA 7 384 well block

Life Technologies Ion Chef instrument (see figure 2.2.8.3.2)

Life Technologies Ion Torrent PGM instrument (see figure 2.2.8.3.2)

Millipore water purification system

Figure 4.2.1.1 (Left) Promega Maxwell DNA Extraction Instrument. (Right) Qiagen EZ1 DNA Extraction Instrument.

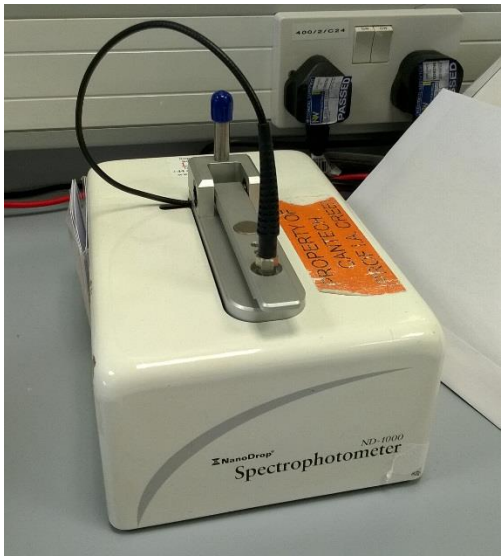

Figure 4.2.1.2 Nanodrop spectrophotometer instrument for DNA quantification.

4.2.2 Kits

Qiagen EZ1 Blood Virus Kit (48 isolations)

Qubit dsDNA High Sensitivity reagent kit

Promega Maxwell Circulating DNA Kit (prototype)

Life Technologies Ion AmpliSeq™ Colon and Lung Panel

Life Technologies Ion Express Barcode 1-16 Kit

Life Technologies Ion Express Barcode 17- 32 Kit

Life Technologies Ion Torrent PGM Sequencing 200 Kit

Life Technologies Ion AmpliSeq™ Library Kit 2.0

Life Technologies Ion PGM IC 200 Kit

Life Technologies Oncomine™ Solid Tumour DNA Kit

4.2.3 Cell Free DNA Analysis

Summary: Patient EDTA blood samples were obtained from the Blood Sciences laboratory and analysed in parallel using both available automated extraction platforms, the Qiagen EZ1 and Promega Maxwell. The former is used routinely for extraction of viral DNA from blood, and the Maxwell was initially designed for FFPE samples. The amount and purity of the cfDNA obtained was then assessed by two quantification methods (Nanodrop and Qubit). Ion Torrent next generation sequencing (NGS) was used to further assess sample DNA quality for NGS library preparation and to determine the mutations present. Finally, cfDNA samples were analysed using TaqMan Array cards to assess the suitability of cfDNA for routine analysis by PCR.

4.2.3.1 Patients and Samples

A series of patients attending the UHCW lung clinic (see table 4.2.3.1) had blood taken as part of their standard clinical investigation. Each patient gave written informed consent for research use of tissue and blood not required for diagnosis. Ethics approval for this was granted to the Arden Tissue Bank by the NRES Committee South Central Ethics Committee (12/SC/0526 – ATBMC-00002). Residual plasma from EDTA samples was obtained from the routine blood samples, the plasma was separated (method below), divided into 1 ml aliquots, and stored at -20°C until extraction. Prior to extraction, samples were thawed at room temperature and then processed according to the manufacturer's instructions, as described in a later section.

4.2.3.2 Plasma Separation

EDTA anti-coagulated blood samples were centrifuged at 1600 x g for 15 mins to separate the plasma from the cellular component. The supernatant was collected and the cellular fraction discarded. The supernatant was centrifuged again at high speed (15,500 x g for 4 minutes) to remove platelets and cellular components such as exosomes. The supernatant was collected and the pellet discarded. If the samples were not to be extracted immediately, the samples were stored at -20 °C until ready to be extracted.

Table 4.2.3.1: Samples and clinical details, n= 40. N.B. PL012 and PL017 excluded due to technical reasons. Unfortunately complete data was not available for all samples. NA = Not applicable.

Plasma Sample No.	Sample type	Age	Tissue diagnosis	Radiological diagnosis	Site of primary tumour	Staging	Clinical comments
PL001	EDTA Blood	61					
PL002	EDTA Blood	59	Clear cell renal cell carcinoma	metastatic disease	kidney	metastatic disease	
PL003	EDTA Blood	Unknown					
PL004	EDTA Blood	52					
PL005	EDTA Blood	42					
PL006	EDTA Blood	63	Metastatic lung adenocarcinoma				
PL007	EDTA Blood	72	NA	RUL, RLL masses			PMH: breast carcinoma - Rx Arimidex. Severe emphysema, lung biopsy not possible
PL008	EDTA Blood	69					
PL009	EDTA Blood	61	NA	benign pleural plaque	NA	NA	
PL010	EDTA Blood	78	Squamous cell carcinoma		lung	T3 N3 M1b	
PL011	EDTA Blood	59	Malignant NET-atypical spindled morphology, IHC in keeping with lung primary		lung	T4 N3 M1b	
PL013	EDTA Blood	59	Malignant NET-atypical spindled morphology, IHC in keeping with lung primary		lung	T4 N3 M1b	
PL014	EDTA Blood	66	Amyloidosis		no tumour	no tumour	PMH: SLE
PL015	EDTA Blood	72					
PL016	EDTA Blood	77	NA		NA	NA	
PL018	Serum	59	Non-small cell carcinoma	LUL mass	lung	T4 N2 M1b	previous lymphoma
PL019	EDTA Blood	73	adenocarcinoma		lung	T1 N0 M1	Bone marrow metastasis
PL020	EDTA Blood	73	adenocarcinoma		lung	T1 N0 M0	Bone marrow metastasis
PL021	EDTA Blood	70	no evidence of malignancy		NA	Cerebral metastases	cerebral metastases
PL022	EDTA Blood	72	Metastatic adenocarcinoma	right lung perihilar mass	lung	T4 N3 M1b	
PL023	EDTA Blood	60	Metastatic small cell carcinoma	Bilateral consolidation	lung	T4 N3 M1b	
PL024	EDTA Blood	60	Metastatic small cell carcinoma	Bilateral consolidation	lung	T4 N3 M1b	
PL025	EDTA Blood	75	Non-small cell carcinoma, favour adenocarcinoma	LLL mass	lung	T4 N2	Initial diagnosis **2011** - Rx carboplatin/alimta and high dose radiotherapy
PL026	EDTA Blood	74	NA	No malignancy - COPD	NA	NA	
PL027	EDTA Blood	76	adenocarcinoma		lung	T4 N2	
PL028	EDTA Blood	86	NA	RUL mass	lung	T2 N0 M0	
PL029	EDTA Blood	76	adenocarcinoma	RUL/RLL consolidation	lung	T4 N2	
PL030	EDTA Blood	76	adenocarcinoma	RUL/RLL consolidation	lung	T4 N3	
PL031	EDTA Blood	70	adenocarcinoma	LUL mass	lung	pT3 pN0	patient had separate tumour <1cm in same lobe
PL032	EDTA Blood	82	NA	LLL mass	lung	T4 N1 M1a	
PL033	EDTA Blood	81	NA	RLL opacity --> resolved	NA	NA	previous Hodgkin lymphoma, breast cancer and tuberculosis
PL034	EDTA Blood	84	NA	Fibrosis	NA	NA	
PL035	EDTA Blood	79	Squamous cell carcinoma	LLL mass	lung	T3 N2 M0	
PL036	EDTA Blood	87	Metastatic squamous cell carcinoma	RUL mass	lung	T3 N2 M1b	
PL037	EDTA Blood	75	adenocarcinoma	LLL mass	lung	pTa N0 M0	
PL038	EDTA Blood	79	adenocarcinoma	RUL mass	lung	pT1a N0 M0	
PL039	EDTA Blood	59	adenocarcinoma (mucinous)	right lung mass	lung	pT4 N0 Mx	
PL040	EDTA Blood	70	Metastatic adenocarcinoma	RUL mass	lung	Not staged	previous Hodgkin lymphoma
PL041	EDTA Blood	78	Metastatic small cell carcinoma	RUL mass	lung	T2a N2 M0	
PL042	EDTA Blood	50	adenocarcinoma	LLL mass	lung	T4 N3 M1b	

4.2.3.3 DNA Extraction from Plasma

Twice centrifuged EDTA plasma samples (procedure described above) were extracted using two different automated extraction platforms: the Qiagen EZ1 instrument (using the Qiagen EZ1 DSP Virus 2.0 Kit) and the Promega Maxwell instrument (using the prototype Promega

Maxwell Circulating DNA Kit). DNA samples were eluted in 60 µl elution volumes for both instruments and stored at -20°C until ready for molecular analysis.

4.2.4 DNA quantification

The quantity of the extracted nucleic acid obtained from EDTA blood was assessed by two methods, listed below:

1. Quantification by Nanodrop spectrophotometer platform.
2. Quantification by the Qubit platform (for method, see section 2.2.7)

Nanodrop procedure: before beginning measurements of samples, the instrument pedestal as cleaned with 1.5 µl dH_2O and a lint free tissue. Two blank measurements were then made with 1.5 µl extraction kit elution buffer to normalise the instrument sensor for background absorbance. To measure the DNA concentration of the extracted DNA samples, a 1.5 µl volume was dispensed onto the Nanodrop instrument pedestal, and the absorbance measured at 260/280 nm. The measurement result is outputted as a value in ng/ml. The instrument was re-blanked with extraction kit elution buffer every five sample measurements, and after all samples had been measured the instrument pedestal was cleaned with dH_2O.

4.2.5 Ion Torrent Next Generation Sequencing of Plasma DNA

Sequencing was performed as previously described (Tops *et al.*, 2015), with the following changes to accommodate Ion Chef loading of the Ion Torrent 318 chips. Stored or newly extracted DNA was checked for content and quality using a Qubit 2.0 Fluorometer. 10 ng of gDNA from each of the samples chosen for NGS analysis was combined with the Ampliseq™ reagents and primer pool for the OncoNetwork 22 gene panel (Tops *et al.*, 2015) and amplified for 30 cycles. After initial amplification the amplified products were partially digested before Ion Express Adapters and Barcode sequences are ligated to the library fragments. Following barcoding the libraries were cleaned up using a magnetic bead method. The cleaned up products were then quantified by the AmpliSeq™ Q-PCR method. Once the libraries had been successfully quantified they were combined and diluted to 50 pM. For Ion Torrent 318 chips, 20 libraries were combined per chip. These library pools were then loaded into the Ion Chef instrument for further library preparation and chip loading. The loaded 318 chips were then run on the Ion Torrent PGM instrument according to manufacturer's instructions.

The Variant Caller plugin (including in the provided Ion Suite software) was used to analyse the aligned sequence data for the identification of hotspot mutations and novel variants. The variant data was filtered using 'High Stringency' parameters to reduce the incidence of false positives (key filters: minimum variant frequency: 3%; minimum quality score: 10; Minimum coverage: 20 reads).

4.2.6 TaqMan Array PCR Assay

As a limited scale pilot study, hotspot positive cfDNA samples were also analysed using a TaqMan Array PCR assay. This assay had been previously validated for detection of 44 common somatic mutations in lung and colorectal FFPE tissue samples (Kikuchi H., 2016). Extracted cfDNA samples were loaded at 50 ng per channel and run according to manufacturers' instructions.

For direct comparison, if a patient sample was found to be positive for a hotspot by NGS on either extract (EZ1/ Maxwell), then both extracts were analysed by the TaqMan array, regardless of whether the other extract had been previously been found to be negative by NGS. In accordance with the previously published validation study, mutant targets producing a Ct below 36.5 were considered positive, any signals coming up later than this were considered negative.

4.2.7 Statistical Analysis

Statistical tests were carried out using R (RC, 2015). Where appropriate, t tests were paired and two sided, with significance level $\alpha = 0.05$. Variances were not assumed to be equal. Pearson correlation coefficients were calculated using rcorr from the package *Hmisc*.

4.3 RESULTS

4.3.1 DNA quantity

The overall DNA recovery from separated plasma samples extracted in parallel both on the Qiagen EZ1 and Promega Maxwell platform is shown in figure 4.3.1.1. The extracts were measured using both Nanodrop and Qubit to determine the isolated DNA concentration. Using the Nanodrop spectrometer method, the Qiagen EZ1 extracts appear to have a significantly higher DNA concentration (mean 41.68 ng/µl) compared to the Maxwell (mean 4.26 ng/µl). This difference is statistically significant ($P<0.01$).

However, when measurements were repeated using the Qubit method, the extracts from Maxwell contained a higher concentration of DNA (mean 2.10 ng/µl) compared to EZ1 (mean 1.04 ng/µl), data shown in figure. 4.3.1.2, although the difference is only marginally above the threshold for statistical significance ($P=0.053$). This suggests that the EZ1 extracts contained a higher level of protein contamination, giving a falsely high reading at the 260/ 280 absorbance frequency. The Maxwell extracts also gave a higher DNA concentration result when measured by Nanodrop (compared to Qubit), however the difference between the results from the two instruments was not as dramatic compared to the EZ1 extracts. This may suggest the Maxwell extracts contained purer DNA samples than the equivalent EZ1 extracts, and that Qubit gives more accurate measurements of DNA content.

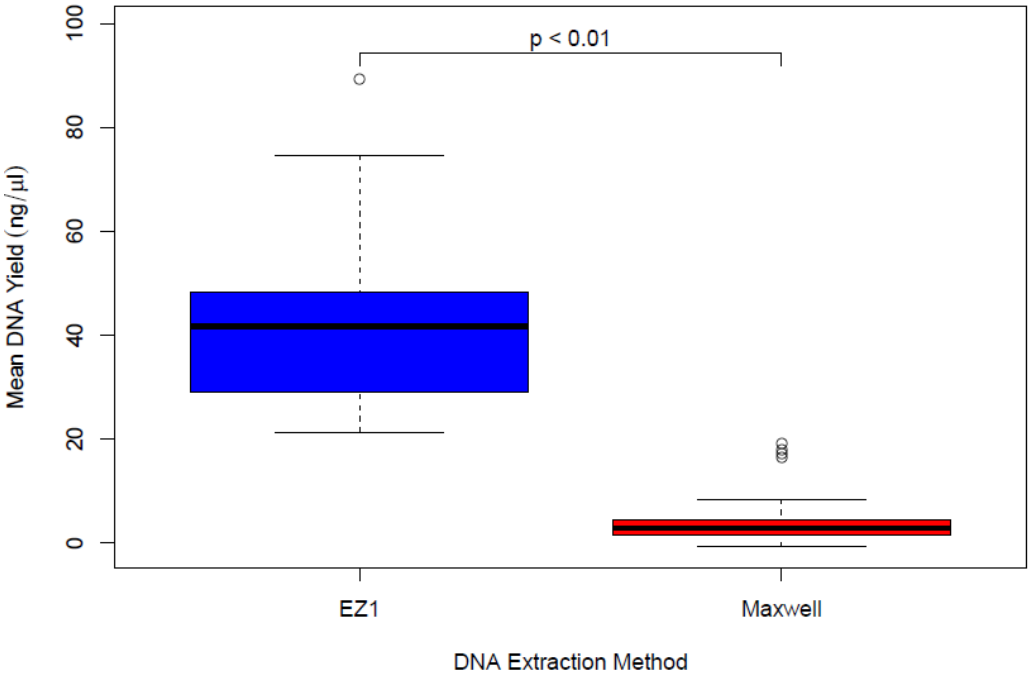

Figure 4.3.1.1: DNA isolation from plasma, using platforms Qiagen EZ1 and Promega Maxwell. Measurements by Nanodrop. Mean values of data group are shown as thick black lines, outlying data points represented as circles.

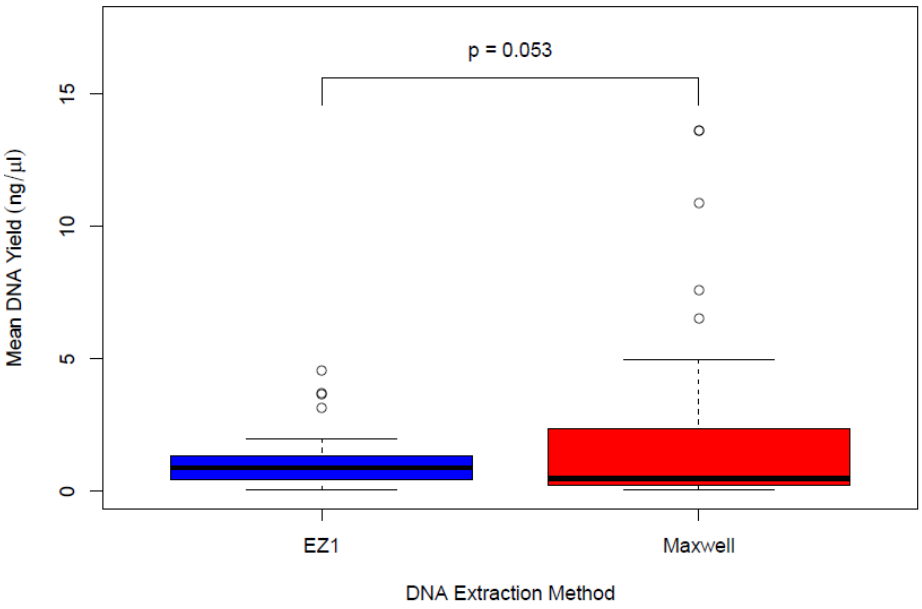

Figure 4.3.1.2: DNA isolation from plasma, using platforms Qiagen EZ1 and Promega Maxwell. Measurements by Qubit. Mean values of data group are shown as thick black lines, outlying data points represented as circles.

The mean cfDNA recovery from the separated plasma samples (measured by Qubit), using both automated extraction platforms is shown in figure 4.3.1.3. The samples are grouped based on the storage time of the original unseparated EDTA blood samples (24, 48 and 72 hours). In all groups, the Promega Maxwell isolated higher cfDNA concentrations than the same samples extracted using the Qiagen EZ1. The highest cfDNA concentrations were found in the 48 hour storage time group. However it is important to point out that none of the these samples were sequentially extracted on concurrent days, and therefore these data do not necessarily reflect DNA degradation over time within the same sample, rather it simply illustrates the high degree of variation of cfDNA levels found in clinical samples.

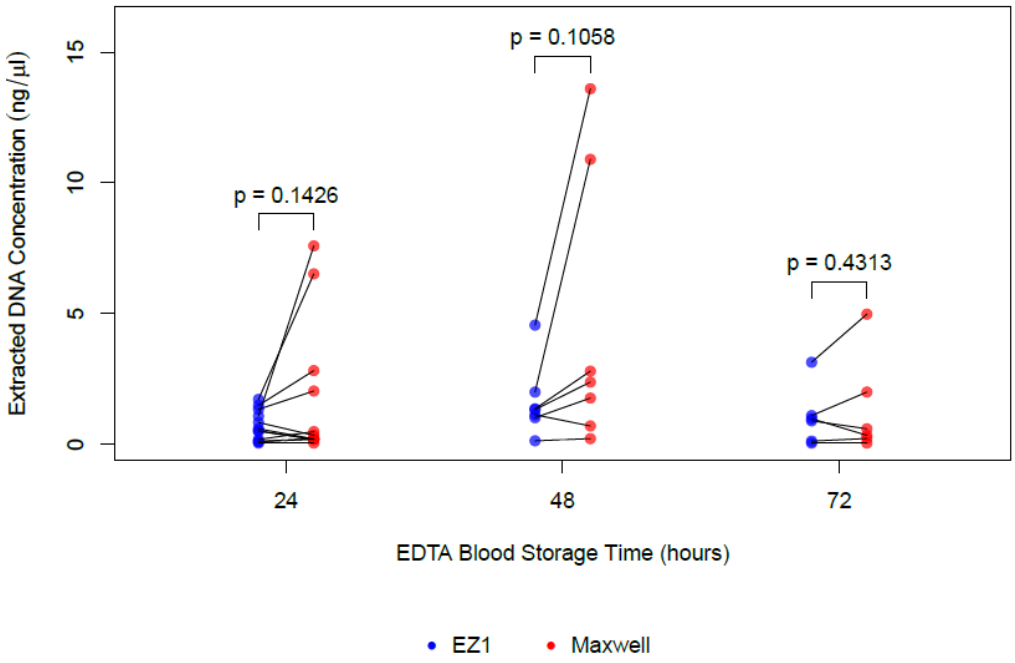

Figure 4.3.1.3: Graph showing the cfDNA recovery from plasma separated at varying storage times. Measurements from Qubit instrument. Black connecting lines indicate paired extracts from the same original plasma sample.

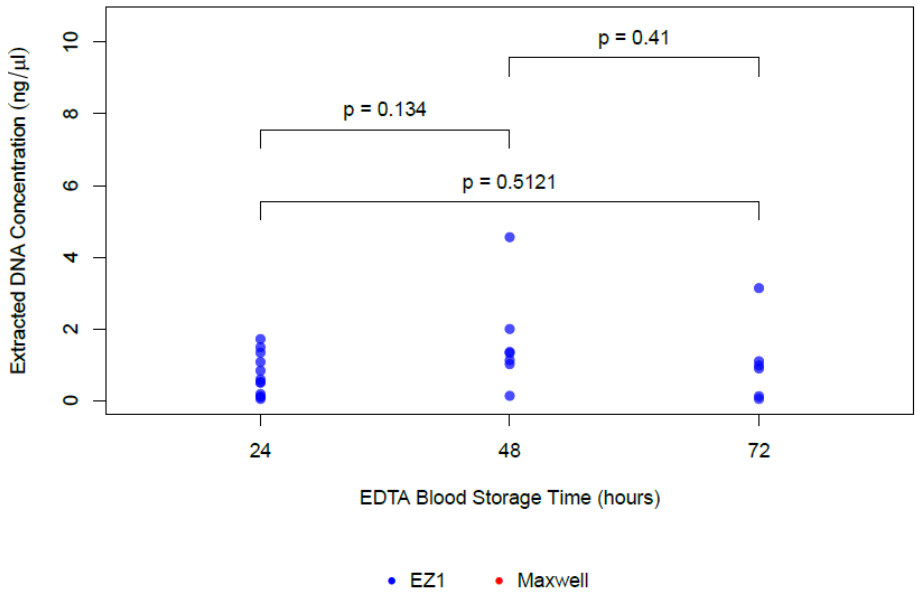

Figure 4.3.1.4: Graph showing the cfDNA recovery from plasma separated at varying storage times. EZ1 platform only. Measurements from Qubit instrument.

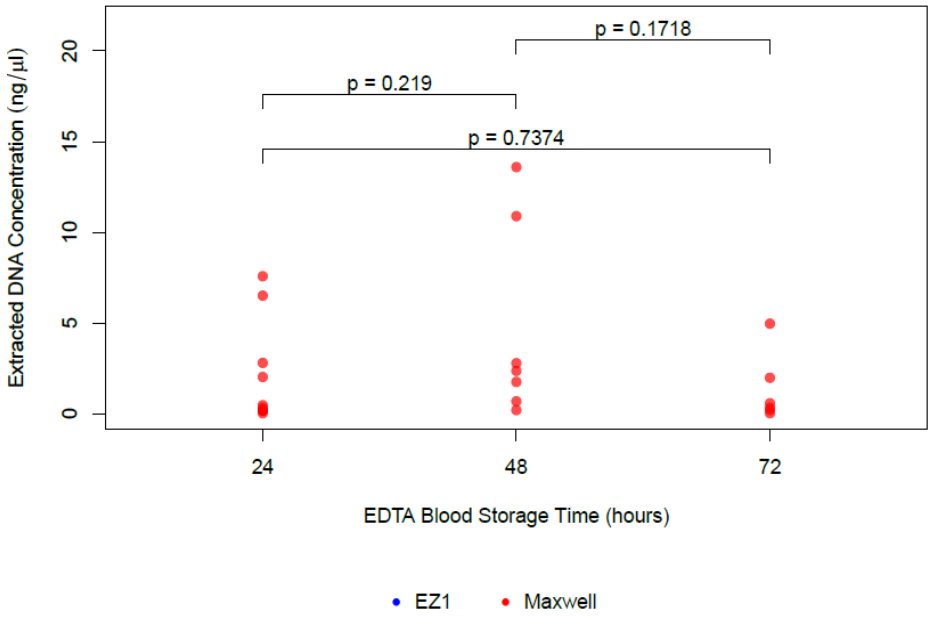

Figure 4.3.1.5: Graph showing the cfDNA recovery from plasma separated at varying storage times. Maxwell platform only. Measurements from Qubit instrument.

Interestingly when the data is separated by EDTA blood storage time, there is no statistical significance between which extraction platform was use or the storage time. The graph illustrating this analysis are shown in figures 4.3.1.4 (EZ1) and 4.3.1.5 (Maxwell).

4.3.2 NGS Library quality: Quantitation EZ1 vs Maxwell

NGS library preparation involved a library quantification step using Q-PCR, using the Ion Torrent Library Quantitation Kit (Thermo Fisher). After completing the quantification step each library had a precise quantification value in pM/ µl, which gives an indication of how efficiently the library preparation process has occurred and reflects the quality of the original template DNA used to create the library. Figure 4.3.2 shows a side by side plot of all the libraries generated using the extracted cfDNA from both platforms. Libraries generated from samples extracted on both platforms are arranged together. The paired samples are plotted in ascending order based on library quantification values derived from the Promega Maxwell extract. The data clearly shows that in a majority of cases the extract from the Promega Maxwell generated a far higher library quantification value (i.e. higher quality library) compared to the equivalent EZ1 extract. There are only four incidences where the EZ1 library has a higher quantification value than the Maxwell library.

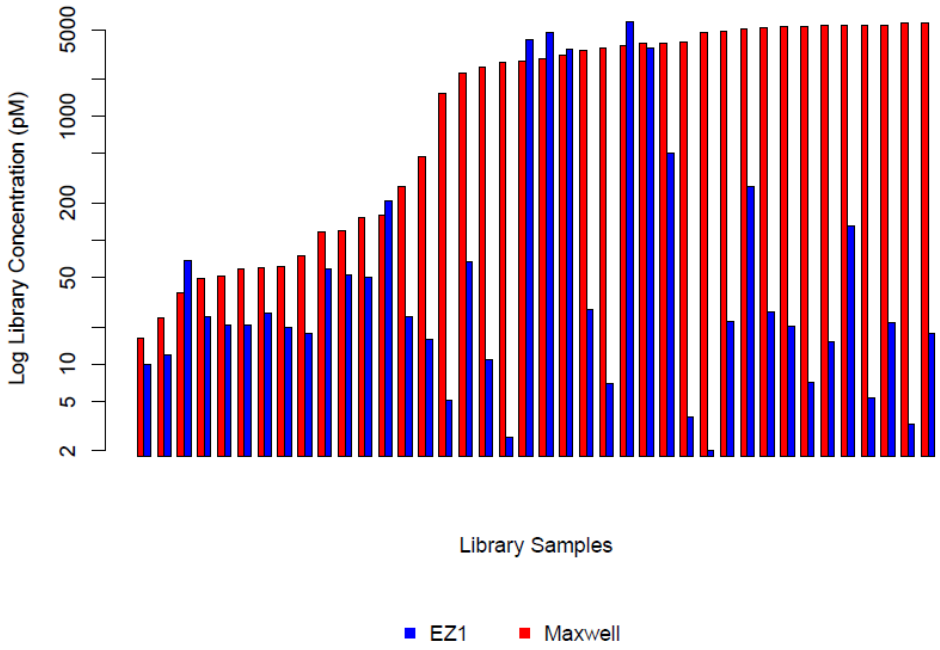

Figure 4.3.2: Graph showing NGS library quantification values, arranged in ascending order from Maxwell extracts. Quantification values from parallel extracted samples are arranged adjacently.

4.3.3 NGS Data Quality

Following the Ion Library Quantitation step all cfDNA libraries were diluted and pooled to 50 pM (where possible) and were processed on the Ion Chef platform for library enrichment and chip loading onto 316v2 or 318v2 chips (Thermo Fisher). The samples on the chips were sequenced using the Ion Torrent PGM instrument according to manufacturer's guidelines. The resulting data was used to analyse both the sequencing success of cfDNA starting material and to compare the relative quality of the extraction methods used to isolate the DNA samples.

Figure 4.3.3.1 shows the original DNA samples grouped by their EDTA storage time, separated by extraction platform, and plots the mean read length of sequencing reads

achieved by these samples. In this case, the data shows that the choice of extraction platform has a statistically significant influence on the mean read length of the NGS results (P<0.01). In all groups, the read lengths achieved by Maxwell extracted DNA samples are longer than those extracted on the EZ1 platform, suggesting that the length of DNA fragments was greater in the Maxwell extracts.

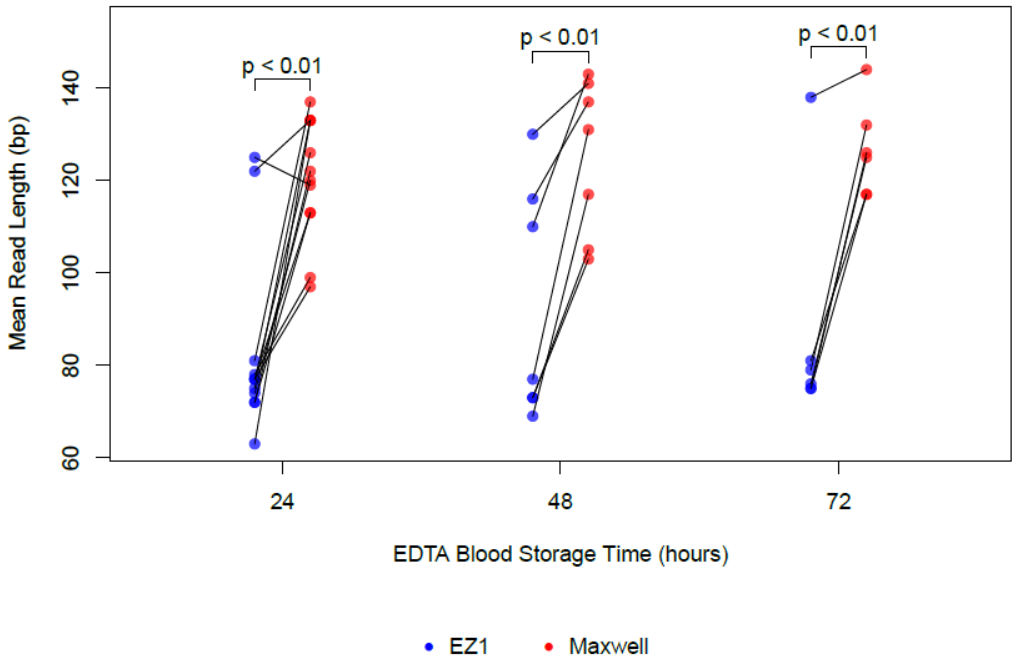

Figure 4.3.3.1: Graph showing EDTA blood storage time against mean read length from NGS analysis. Samples are grouped by storage time and separated by extraction method. Black connecting lines indicate data from paired extracts from the same original plasma sample.

The mean read length data for the individual extraction platforms are shown in figures 4.3.3.2 (EZ1) and 4.3.3.3 (Maxwell). The data appears to show that EDTA storage time (from 24 to 72 hours) does not have a significant influence on the mean read length. Mean read length results are consistent across all samples groups (based on storage time).

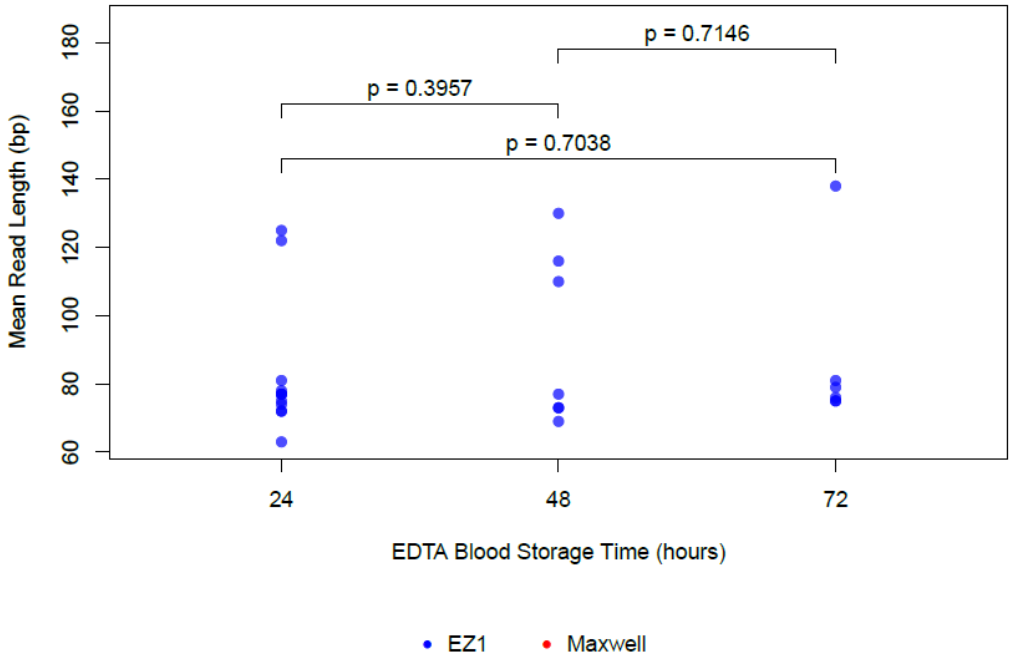

Figure 4.3.3.2: Graph showing EDTA blood storage time against mean read length from NGS analysis. EZ1 platform only.

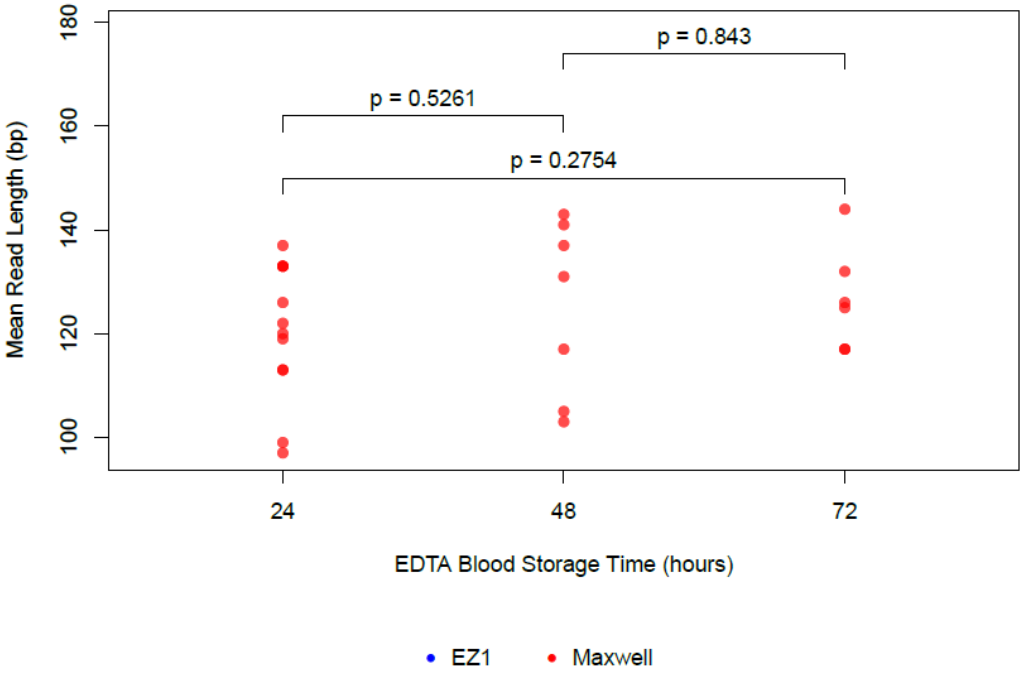

Figure 4.3.3.3: Graph showing EDTA blood storage time against mean read length from NGS analysis. Maxwell platform only.

Figure 4.3.3.4 shows the DNA samples grouped by their EDTA storage time, separated by extraction platform, and plots the mean number of sequencing reads achieved by these samples. This plot shows that in the 24 and 72 hour groups, the choice of extraction platform has a statistically significant difference on the number of sequencing reads generated. At 48 hours there is no significant difference between the extraction platforms.

Libraries from Maxwell extracted samples show increasing variability of read number as EDTA storage time increases, with the lowest and most variable number of reads occurring in the 72 hour group (range 61176- 563406 reads), suggesting increased fragmentation at that point. The EZ1 extracted samples demonstrated a high degree of variability, and mean read numbers were not consistent across the storage time groups, indicating a lack of consistency in the quality of cfDNA isolated by this platform. The data from figure 4.3.3.4 appears very similar to the cfDNA recovery data from figure 4.3.1.3, so this logically suggests initial cfDNA concentration has an influence on NGS read numbers when sequenced.

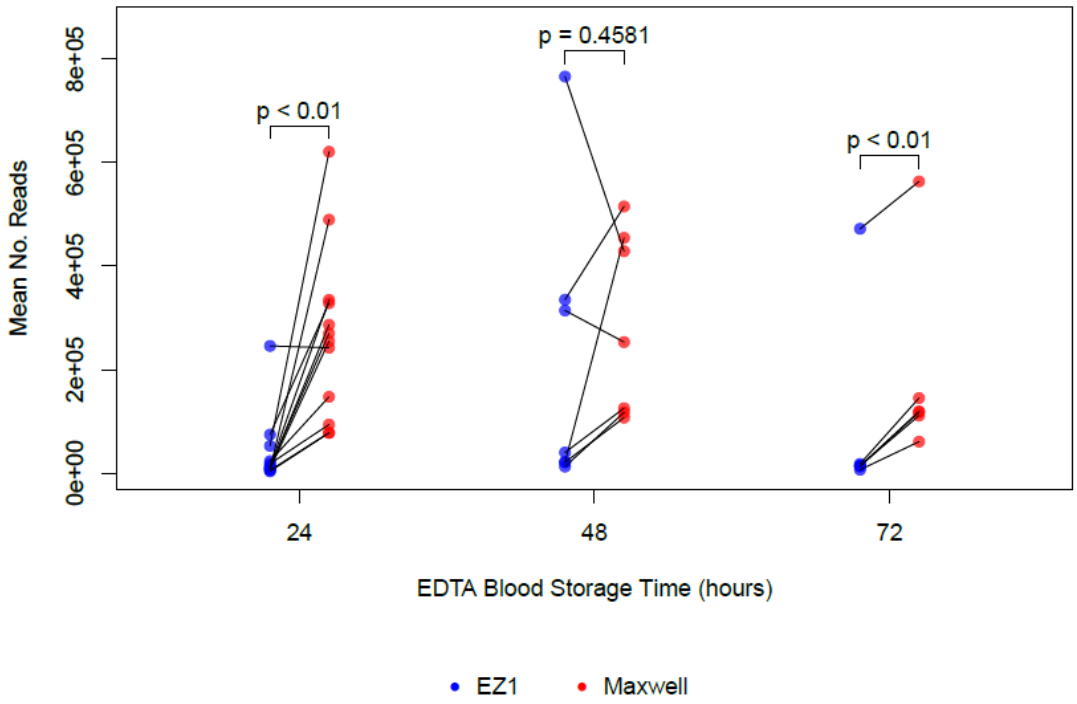

Figure 4.3.3.4: Graph showing EDTA blood storage time against mean number of reads from NGS analysis. Samples are grouped by storage time and separated by extraction method. Black connecting lines indicate data from paired extracts from the same original plasma sample.

When the read numbers data is separated by extraction platform, the data again shows that EDTA storage time does not have a significant effect on this aspect of the NGS results. See figures 4.3.3.5 (EZ1) and 4.3.3.6 (Maxwell).

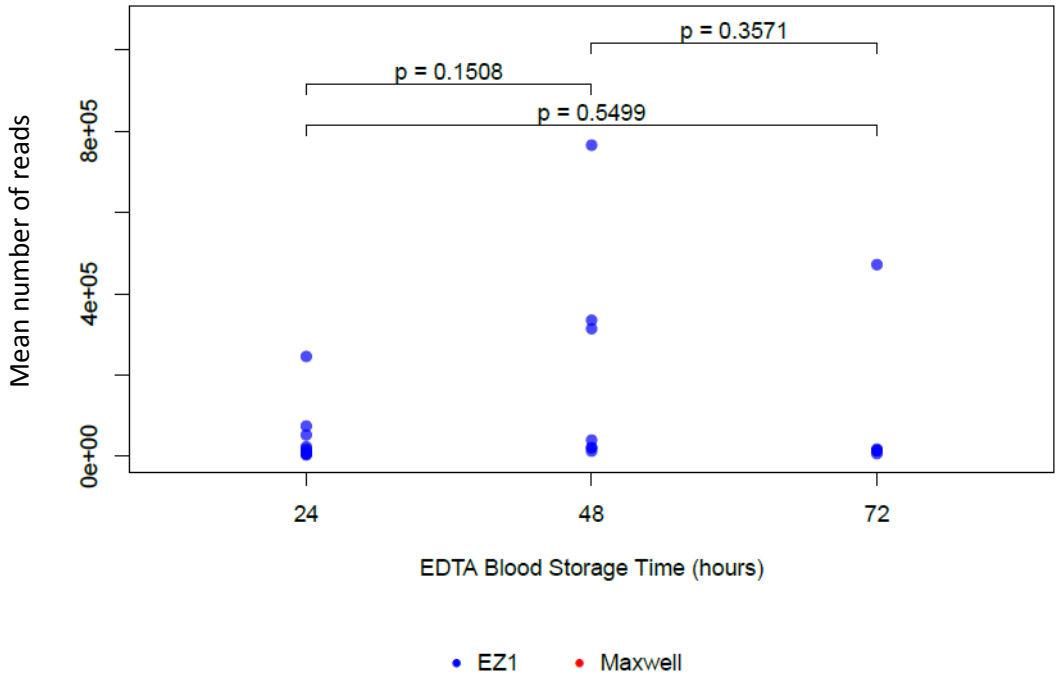

Figure 4.3.3.5: Graph showing EDTA blood storage time against mean number of reads from NGS analysis. EZ1 Platform only.

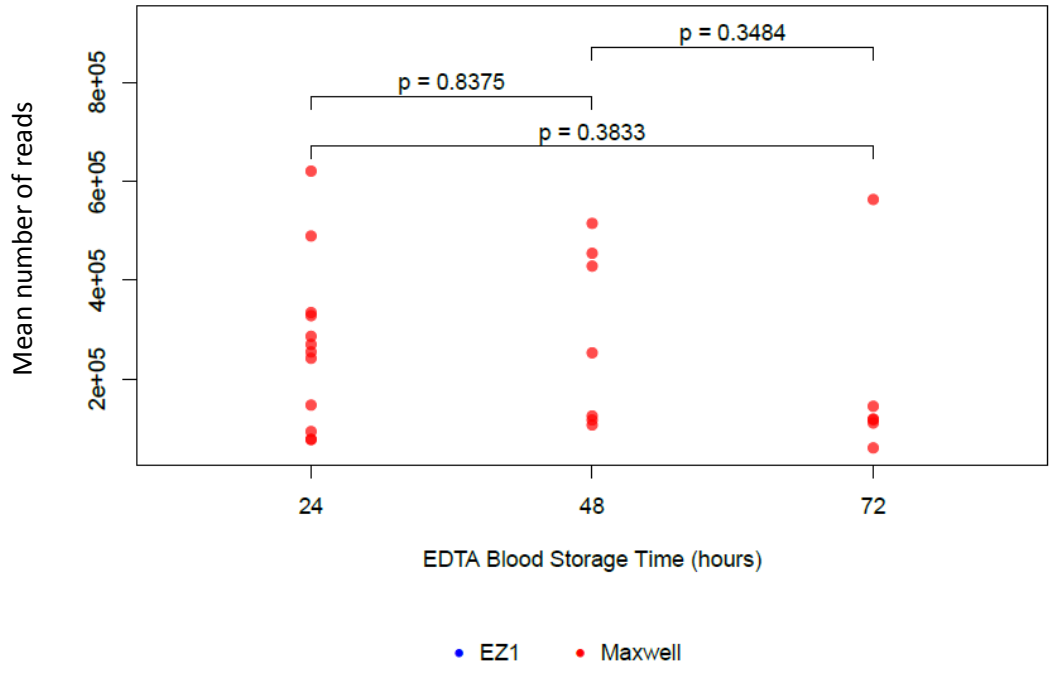

Figure 4.3.3.6: Graph showing EDTA blood storage time against mean number of reads from NGS analysis. Maxwell Platform only.

Figure 4.3.3.7 shows the correlation between read length and coverage achieved by the two sets of libraries (Maxwell and EZ1 extracted). Libraries that failed sequence analysis (i.e. data was of insufficient quality to perform coverage or variant caller analysis) were removed. With the exception for four outliers, the Maxwell libraries demonstrate a consistent grouping of read length and coverage. By contrast EZ1 libraries are much less consistent, with read length generally being much lower and associated coverage highly variable. Centroid markers for the two datasets show that the Maxwell extracts have generally higher read length, whereas the two groups are equivalent in terms of coverage (number of reads per amplicon).

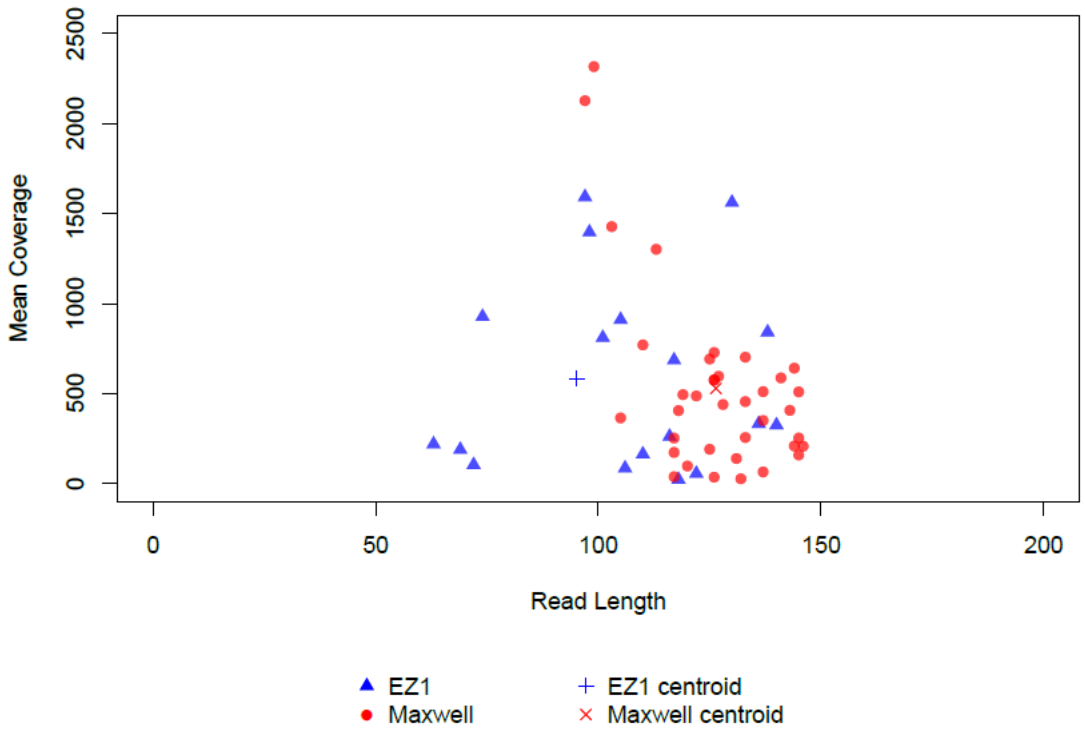

Figure 4.3.3.7: Graph correlating read length against coverage from NGS analysis. Samples are grouped by extraction method.

4.3.4 NGS Variant Detection

Library samples sequenced on the Ion Torrent PGM instrument (that generated data of a suitable quality) were analysed by the Variant Caller plugin included in the Torrent Suite server software. Sample data was analysed using the pre-set 'High Stringency' analysis parameters to filter out low quality data and low frequency variants to reduce the likelihood of false positives in our dataset (filters described in Methods section 4.2.5).

Variant Caller analysis identified a wide range of novel variants and hotspot mutations detected in our cfDNA samples. Figure 4.3.4 shows the number of variants (novel and hotspots) detected in all the paired samples extracted on the two extraction platforms. In all but four samples, more variants are detected in the libraries from Maxwell extracts than the equivalent samples from the EZ1 platform. This suggests that the Maxwell extracts contained higher quality (higher purity and less fragmented) DNA which has had a direct influence on the quality of the NGS data generated downstream.

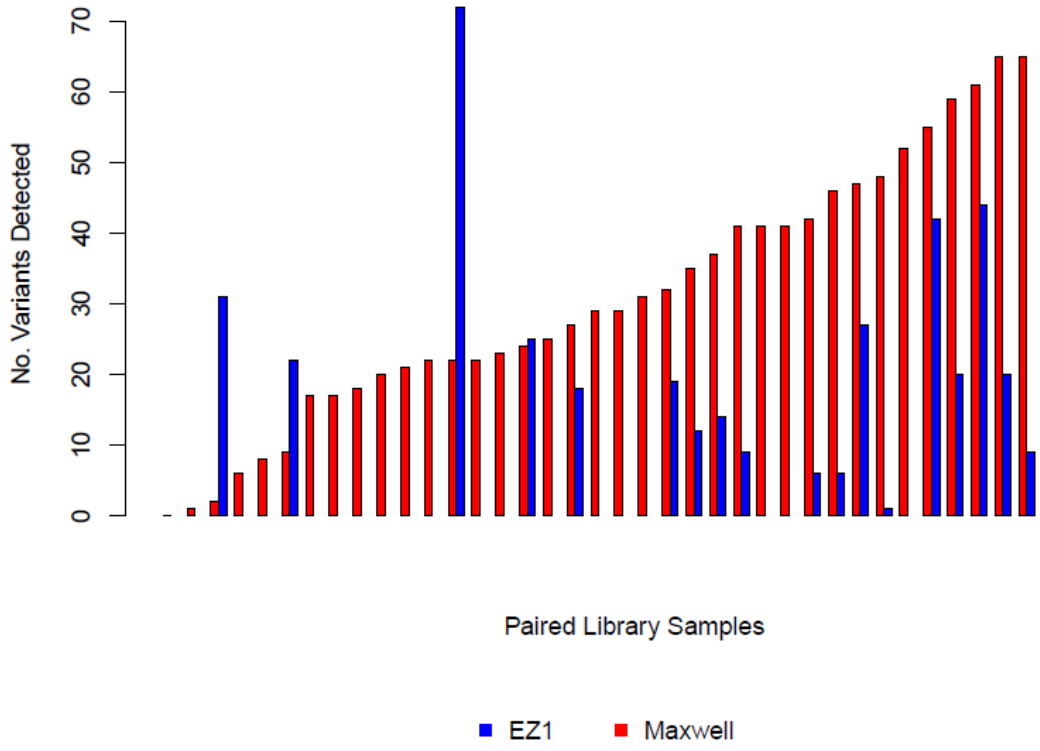

Figure 4.3.4: Graph showing the number of variants detected in each cfDNA sample, arranged in ascending order (based on Maxwell extract variant detection). Samples are grouped with paired EZ1 extracts.

4.3.5 NGS Hotspot Detection

The high stringency analysis parameters used by the Variant Caller plugin produced data which is likely to be free of false positives, so the assumption was made that any hotspots remaining after this filtering can be confidently called positive. Figure 4.3.5 shows the numbers of hotspot mutations identified in our cfDNA samples. In total eight extracts tested positive for a total of 12 hotspot mutations. These eight positive extracts were derived from six original clinical samples. Of the 12 hotspot mutations identified, a total of nine were identified in the Maxwell extracted samples, whereas only three were detected in the EZ1 extracted samples.

Figure 4.3.5: Graph of hotspot mutations detected in cfDNA libraries. Libraries that could not be analysed are designed with a coloured dot in the NA (not applicable) column.

There was very little concordance between the results from different extraction platforms. Of the six clinical samples that tested positive for hotspot mutations, the results from the Maxwell and EZ1 instruments only agreed in one case (PL042). In three of the other cases (PL027, PL039 and PL041), hotspots were detected in the Maxwell extracts whereas none were detected in the corresponding EZ1 extracts. In another case (PL021), two hotspots were detected in the Maxwell extract whereas only one was detected in the EZ1 extract; and finally in one case (PL008) a single mutation was detected in the EZ1 extract whilst the corresponding Maxwell extract was negative for hotspots.

Table 4.3.5: Hotspots detected from cfDNA samples. Hotspots detected in EZ1 extracts are highlighted in blue. SNP = Single Nucleotide Polymorphism, DEL = Deletion.

Sample Name	Platform	Frequency	Quality	Type	Allele Name	Gene ID	Mutation Details	Coverage	Allele Cov	Strand Bias
PL021	MAXWELL	21.7	2145.53	SNP	COSM516	KRAS	p.G12C	1659	360	0.50
PL021	EZ1	22.9	122.64	SNP	COSM516	KRAS	p.G12C	83	19	0.57
PL021	MAXWELL	17.1	1417.11	SNP	COSM20957	STK11	p.D194V	1651	283	0.55
PL008	EZ1	7	50.75	SNP	p.Gly774Glu	DDR2	p.Gly774Glu	316	22	0.95
PL039	MAXWELL	10.2	37.68	SNP	COSM10645	TP53	p.C176F	88	9	0.56
PL039	MAXWELL	5.9	9	DEL	COSM20871	STK11	p.P281fs*6	118	7	0.58
PL039	MAXWELL	5.9	9	DEL	COSM12924	STK11	p.P281fs*6	118	7	0.58
PL041	MAXWELL	50.8	1293.7	SNP	COSM706	MET	p.E168D	246	125	0.50
PL041	MAXWELL	48.5	148.14	SNP	COSM710	MET	No COSMIC record	33	16	0.55
PL042	MAXWELL	40.2	873.28	SNP	COSM775	PIK3CA	p.H1047R	241	97	0.57
PL042	EZ1	43.4	407.39	SNP	COSM775	PIK3CA	p.H1047R	106	46	0.60
PL027	MAXWELL	52.5	10867.4	SNP	COSM706	MET	p.E168D	1986	1042	0.51

Table 4.3.5 shows a detailed overview of the hotspot mutations detected in our sample set. Quality scores and coverage were highly variable amongst the hotspots identified, which is most likely related to individual properties of the template DNA (concentration, fragmentation). However all these hotspots were retained after high stringency filtering by the Variant Caller plugin in the Torrent Suite software, and therefore these are likely to be genuine mutations from the clinical samples. Of most interest is the detection of the actionable *KRAS* activating mutation (COSM516, *KRAS* c.34G>T p.Gly12Cys) detected in sample PL021, and this mutation was detected using both the Maxwell and EZ1 extracted DNA, although the quality score and coverage are far higher for the Maxwell sample.

4.3.6 Performance of cfDNA in TaqMan Array Assay

Six sets of paired samples (which tested positive for hotspot mutations by NGS) were analysed in parallel by TaqMan array. The results are shown in table 4.3.6.The samples were analysed in sets of six over two runs. Unfortunately the second run failed so the final six samples could not be analysed (data not shown). For the six extracts that could be analysed,

the TaqMan Array assay gave 100% concordance with the NGS results. Clinical data was only available for two of the samples analysed. Samples 1 and 3 (01 MXW/ 01EZ1 and 03 MXW/ 03 EZ1) were metastatic adenocarcinomas. Staging information was only available for sample 1, which was staged at T4 N3 M1b. Considering the advanced stage of the disease, the presence of a detectable hotspot mutation is not surprising. No clinical data was available for sample 2 (02 MXW/ 02 EZ1).

Table 4.3.6: Results from TaqMan Array (TA) analysis of paired CFDNA samples (positive for mutations on NGS).

Sample no.	DNA input (ng)	PCR PASS/ FAIL	Positive Targets	CT	TA vs NGS	Disease type	Disease Stage
01 MXW	50.00	PASS	KRAS G12C	31.458	MATCH	Metastatic adenocarcinoma	T4 N3 M1b
01 EZ1	50.00	PASS	KRAS G12C	33.042	MATCH	Metastatic adenocarcinoma	T4 N3 M1b
02 MXW	4.10	PASS	None detected		MATCH	Not available	Not available
02 EZ1	29.47	PASS	None detected		MATCH	Not available	Not available
03 MXW	11.22	PASS	None detected		MATCH	Metastatic adenocarcinoma	Not staged
03 EZ1	7.07	PASS	None detected		MATCH	Metastatic adenocarcinoma	Not staged

4.3.7 Disease Stage versus cfDNA Levels in Plasma

Clinical background information was available for a majority of the plasma samples used in this study. Where staging information was available, the cfDNA concentrated from the corresponding plasma has been recorded. Figure 4.3.7 shows the cfDNA levels extracted from plasma against the staging information for that patient.

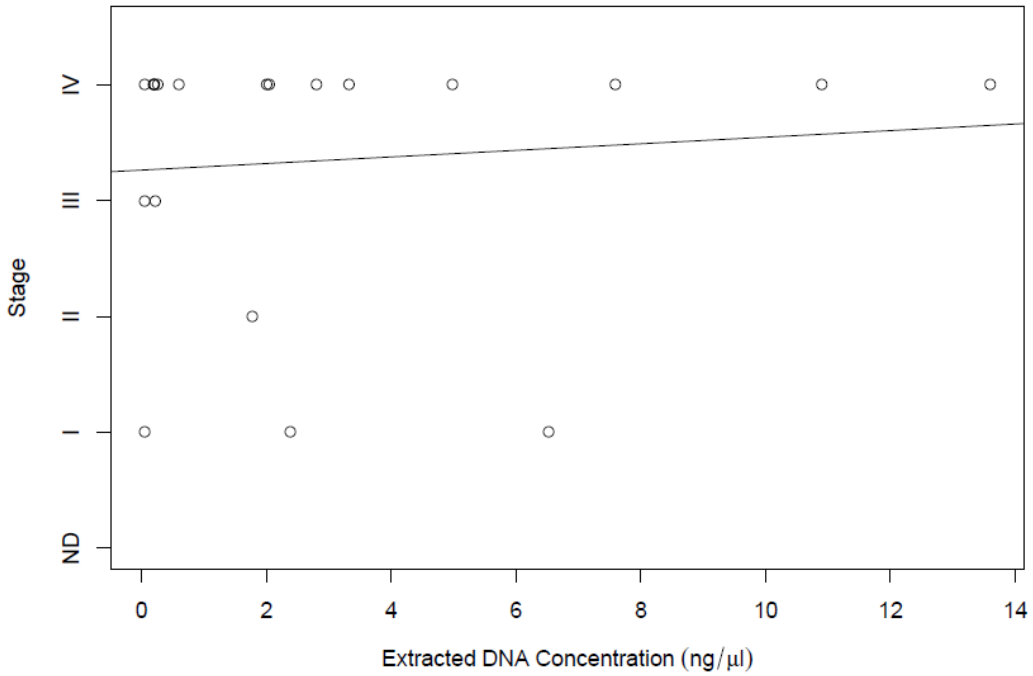

Figure 4.3.7: Graph showing cfDNA concentrated against patient disease stage. Trend line shows a majority of samples in stage 4. ND = No Diagnosis.

From this dataset, there is no significance between the cfDNA levels at different stages. A basic observation from the figure shows that there is a broader range of cfDNA levels at stage IV, but a majority of samples are at the lower end of the DNA concentration scale.

4.4 DISCUSSION

In this investigation I have found clear evidence that both the extraction platforms involved in this study are able to isolate cell free DNA from plasma, although clearly one platform produced better quality DNA than the other.

The first research question in this study was: **can cfDNA be extracted from standard EDTA blood samples, using standard equipment used in the diagnostic laboratory?** Our data has shown that standard EDTA blood samples, obtained through the routine blood sciences services, can be used to obtain cfDNA samples using routine equipment. Whilst these EDTA blood samples were sub-optimally stored (according to other literature) it was reassuring to discover that cfDNA can still be recovered from EDTA blood that has been stored at 4°C for 72 hours after collection. In crude terms mutation detection in this context simply requires a positive or negative result, so therefore sub-optimally stored samples are sufficient for giving this sort of data output. Of course the ideal conditions would use optimally stored cfDNA, however the results here suggest there is some robustness in the methods I have employed.

The discrepancy between the DNA concentration values obtained on Nanodrop and Qubit suggest that the EZ1 extracts contained a significant level of protein contamination, giving a falsely high reading at the 260 nm absorbance frequency. This phenomenon has been described in other studies (Kapp *et al.*, 2015; Simbolo *et al.*, 2013). The Maxwell extracts were more consistent when measured by the different quantification methods. This may suggest the Maxwell extracts contained purer DNA samples than the equivalent EZ1 extracts.

This slight increase is cfDNA in the 48 hour group may be due to the shedding of DNA from lysing exosomes and white blood cells in the blood whilst being stored. This could therefore imply that EDTA blood samples stored for longer than 24 hours have an increased risk of contamination of the plasma by non-cell free DNA being shed during storage.

The data appears to show that EDTA storage time (from 24 to 72 hours) does not have a significant influence on the mean read length or the number of reads (coverage). This seems contrary to other publications where cfDNA is described as rapidly degrading and that the results of downstream analysis would be negatively affected by extended periods of EDTA storage before plasma separation. It is important to state however that whilst read length and number of reads are crucial factors effecting NGS data quality, they are not the only factors to consider. It seems that cfDNA (whilst rapidly metabolised in the body) seems fairly stable at 4˚C in EDTA blood (at least up to 72 hours). The data from figure 4.3.3.6 (storage time vs. number of NGS reads) appears very similar to the cfDNA recovery data from figure 4.3.1.5, so this suggests cfDNA concentration has a significant influence on NGS coverage when sequenced.

It was unfortunate that staging information was not available for more patients. Typically, cfDNA concentration would theoretically be expected to increase with disease progression, but as a majority the samples in our dataset were in stage IV (or unknown stage) then definitive conclusions cannot be drawn from the data.

The second research question in this study was: **of the available standard extraction platforms, which is most efficient at extracting cfDNA, and which yields the best quality DNA for downstream applications?** The Promega Maxwell instrument using the prototype cell free DNA extraction kit was the superior platform in this study. In the majority of cases,

DNA from this instrument was of higher yield and quality, and gave consistently better results when used for NGS analysis. In terms of efficiency, it is important to remember that the sample input volume for the EZ1 platform was only 400 µl, compared to 1 ml for the Maxwell. Given that the Maxwell can accept 2.5 times the plasma volume of the EZ1, it would be logical to expect that the cfDNA yield from the Maxwell would be approximately 2.5 times higher than that of the EZ1. As the average Maxwell extract was less than 2.5 times the EZ1 average, this might suggest that the Maxwell is slightly less efficient than the EZ1. The key difference however is clearly the DNA quality. In crude terms the EZ1 elutes more DNA per µl of plasma input, but the quality is inferior, shown by the Qubit quantification results, and the NGS analytics. This suggests the Maxwell extraction procedure excludes contaminants and highly fragmented DNA molecules.

Probably the clearest demonstration of the effect of extraction platform on NGS data quality is shown in figure 4.3.3.1 and figure 4.3.3.4. In these figures there is a clear statistically significant difference between the relative performances of the extraction platforms, with the only exception being coverage at the 48 hour storage time (figure 4.3.3.4). In all other cases, the libraries generated from Maxwell extracts have longer read length and coverage than the equivalent EZ1 extracts (see figure 4.3.3.7). This is again most likely related to the purity and fragmentation of the eluted DNA samples produced by these instruments. Another consideration in this data is contamination of the extracted plasma DNA by genomic DNA from lysed white blood cells. Due to the extending storage times, it is likely this may have occurred in a proportion of the samples, and may account for some of the longer read length amplicons shown in the NGS data. Unfortunately given the limitations of our sample collection and laboratory set up it was not possible to control/ mitigate this process.

The Taqman array results, whilst limited, give an insight into the relative quality of the DNA samples isolated from the original plasma. Whilst there are a number of NGS analytics showing the Maxwell extracts the be superior, all DNA samples analysed here have given positive results in the TaqMan Array control assays, in some cases with very low DNA input (table 4.3.6). This is most likely a combination of the robustness of the TaqMan Array assay and the quality of the DNA samples. This data is limited however as the number of samples tested is low. It is encouraging however that all the TaqMan Array results matched those determined by NGS, albeit a small number.

The final research question in this study was: **Can cfDNA be used to detect cancer mutations, using pre-designed NGS assays?** Our data shows that cfDNA is a suitable analyte for the Ion Torrent AmpliSeq™ NGS assay designed for use with FFPE tissue, assuming the DNA is isolated using a platform and kit specifically designed for cfDNA. The Ion Torrent AmpliSeq™ kits are of course designed for use with fragmented DNA derived from FFPE tissue blocks, and are robust enough for this use. Others have examined clinical series and find similar results (Garrigou *et al.*, 2016; Caen *et al.*, 2015; Vanni *et al.*, 2015). Library samples generated from cfDNA samples generally gave good results at all quality control steps and generated good quality sequence data.

Under the conditions used, the Maxwell appears to have greater consistency than the EZ1 which may reflect their kit and equipment design. As previous emphasised, the Qiagen Virus Kit is not designed for isolation of cfDNA from plasma, so this needs to be taken into consideration when judging the results. Despite this, it was still possible to isolate cfDNA with potential diagnostic utility from plasma using the EZ1 and use the extracted DNA for molecular methods (PCR and NGS) however the consistency and quality of data was inferior

to that of the extracts from the Maxwell platform. It is important to note that other manufactures (e.g. Siemens, Perkin-Elmer and others) have instruments of their own with varying throughput, plasma volumes etc. for the isolation of cfDNA. To our knowledge, no other comparisons between these systems have been done, and this is an omission making it difficult for diagnostic laboratories to invest in new equipment for cfDNA analysis. Qiagen currently offers a circulating nucleic acid kit for manual extraction, but they have yet to develop a similar kit for their automated platforms.

Results obtained using the Ion Torrent PGM instrument with cfDNA are encouraging as a wide range of novel variants and hotspot mutations were successfully identified, and the data was generally of high quality (see table 4.3.5). This data suggest that cfDNA can be used for detection of actionable cancer mutations.

The limited number of samples tested using the relatively new TaqMan Array technology were very encouraging, as this suggests cfDNA can be used in this far cheaper assay for mutation screening, without the need to perform the full NGS procedure for all cases. The NGS workflow is expensive and time consuming, so the TaqMan Array assay offers a far more rapid and economical testing format. The main limitations of the TaqMan Array assay were the relatively small range of mutations it can detect (44 mutations across *EGFR*, *KRAS*, *NRAS* and *BRAF*) compared to the NGS, and it's relatively high DNA requirement (50 ng for TaqMan Array, 10 ng for NGS). It was impressive to observe however that even with a very low DNA input (4- 11 ng) the TaqMan Array control assays still gave a positive signal, demonstrating that this assay has a high tolerance for low DNA input below the recommended amount. This limitation could be negated by improved cfDNA isolation systems, or the ability to extract cfDNA from larger volumes of plasma, therefore increasing yield for downstream applications.

It is important to take into consideration the limitations of this study when considering the potential for future work. The technical limitations (non-equivalent extraction kits) have been explicitly described, and therefore the performance of the EZ1 platform cannot be expected to be comparable to the Maxwell instrument. This however does illustrate the point that it is important to use appropriate kits for specific applications, especially for work with clinical / diagnostic implications. Whilst a manual extraction kit is available from Qiagen, these sorts of kits are difficult adopt in the diagnostic setting, therefore this study has focused only on automated methods. Automated methods are rapid, consistent and have reduced hands on time compared to manual methods, and therefore they are far better suited to the diagnostic setting.

Another major limitation was access to samples and the associated sub-optimal EDTA storage time, most likely resulting in a potential reduction in sample quality at the time of plasma separation. It was surprising to observe that cfDNA derived from EDTA stored for 72 hours before separation could still be used for NGS analysis, however it is likely data quality would have been improved had the sample been available earlier.

Chapter Five – Discussion

5.1 INTRODUCTION

This project aimed to research, develop and validate new approaches to detecting clinically actionable oncogenic mutations in patient tumour DNA, with the aim of identifying potential clinically useful methods to improve early diagnosis and companion diagnostics. The current status of *EGFR* and *KRAS* mutation detection methods was reviewed in Chapter One, showing that at the beginning of the study (c.2013/14) this process is almost entirely reliant on FFPE tissue and PCR based testing methods, and this is largely still the case at time of submission (c.2017). There was however a considerable interest in exploring new sources of tumour DNA, most notable cfDNA and ctDNA, although work on these analytes was in its early stages at the beginning of the project. In order to follow this area of research, and number of different approaches were adopted in this study.

5.2 SINGLE GENE TESTING

The first approach was to follow the established method of single gene testing with a molecular assay. This involved assisting in the development of the prototype PNA PCR assay (GeneFirst MMD EGFR PCR and MMD KRAS PCR kits), described in Chapter Two. There was a lot of supporting evidence from the literature that this was a logical approach due to the numerous PCR-based assays in use diagnostically (High Resolution Melt assays, Qiagen Therascreen, Life Technologies castPCR for *KRAS*, Roche Diagnostics COBAS test for *BRAF*) and a number of other PCR techniques in development/ validation (digital PCR, SNaPshot). PCR had the cost advantage over the more expensive and time consuming NGS assays. Unfortunately the prototype GeneFirst PNA PCR assays displayed a very low level of

specificity resulting in very frequent false positive results, along with a certain degree of assays failure on the controls. With any assay development, a certain level of discrepancy or variations in reliability are to be expected at early stages, however the fundamental problems in the assays continued throughout all five successive versions. This eventually led to this approach to be discontinued in early 2015, and further development work by GeneFirst was subsequently abandoned.

Beyond the highly variable sensitivity, specificity and reliability issues, the assay was also time consuming and complicated to set up due to its multiple master mixes (and multiple replicates required per sample). It is worth noting however that this is not specific to the GeneFirst assays, as many other diagnostic PCR assays also employ multiple master mixes and require multiple replicates per sample to establish a reliable average signal (Ct) for mutation calling. The consequence of these assays using multiple master mixes and replicates resulted in a very high sample volume and DNA requirement, which for some sample types (in particular small cytology samples) resulted in very low DNA input for the individual reactions, thus creating the possibility of assay failure or false negative results due to the lack of template. The complicated set up also meant the GeneFirst assays offered no advantage over other assays in terms of easy-of-use or hands-on-time.

The GeneFirst assays were tested to a very limited extent with cfDNA (see section 2.3.12) and the results were consistent with the variable results found when using tissue. The assays were designed with the intention to apply them to cfDNA testing, however the results showed them to be unsuitable for both testing tissue and cfDNA.

Whilst PNA PCR has been shown to be successful at detecting low concentration targets in a number of other settings, for example the CE-IVD marked PNAClamp™ EGFR Mutation

Detection kit (Panagene, Daejeon, Korea) in this case unfortunately the assays could not be validated for clinical use. The findings of this phase of the study do not eliminate PNA PCR based methods as a suitable methods for detecting oncogenic mutations, but it does highlight the difficulties involved in assay development. This work demonstrates that assays designed for diagnostic use need to be well designed in order ensure both set up and result interpretation are efficient and as simple as possible.

The clinical implications of false positive and false negative results in NSCLC or CRC are severe, and as a result high sensitivity and specificity are essential for any diagnostic assay that effect clinical decisions. In NSCLC, the presence of activating *EGFR* mutations is an indicator that a patient is suitable for anti-EGFR antibody therapy. A false positive result would result in the patient being given drugs that would be of no therapeutic benefit instead of a more appropriate treatment. A false negative would result in an activating mutation not being detected, resulting in the patient not receiving the anti-EGFR therapy, and most likely receiving an alternative which is both ineffective and likely to be toxic. In the case of CRC, the situation is different. The presence of activating *KRAS* mutations indicates that anti-EGFR therapy will be ineffective, therefore in the case of a *KRAS* positive CRC sample then the patient will be given an alternative treatment to anti-EGFR therapy.

The final issue related to this approach is that single gene testing is not the optimal method for routine testing of clinical samples. Clinically actionable mutations in lung and colorectal cancer (and many other cancers) often occur in multiple genes, and therefore it is logical to perform genetic analysis on multiple genes when analysing a clinical sample. Performing separate assays for mutations in a single gene has limited usefulness and is very inefficient in terms of staff time, and also demands a larger amount of genetic material for analysis, the availability of which can be very varied in certain contexts.

5.3 MULTIPLE GENE TESTING

After the unsuccessful development of the GeneFirst PNA PCR and the single gene testing approach, the second phase of the project was to move to multiple gene testing. This involved the use of an alternative PCR format- TaqMan Array, covered in Chapter Three. The layout of this plate was devised by searching the most common oncogenic mutations on COSMIC. A series of lung, colorectal and melanoma samples were extracted and analysed using the TaqMan Array assay. The assay performed extremely well, matching all the previously obtained results (from the Therascreen, castPCR or COBAS assays) except in a small number of cases. These discrepancies were analysed by NGS (if not already done so) in order to ascertain the true genotype.

The dramatically superior performance of this assay compared to the GeneFirst assays was due to a number of factors. Firstly, the TaqMan Array assays are based around the already well proven castPCR technology, which has been shown to be both robust and reliable (Bolton *et al.*, 2015) and a version of this was previously used in UHCW Pathology for its routine cancer gene mutation detection for colorectal samples. Reformatting the castPCR assays from a 96 well format (20 µl per reaction) to an enclosed 384 well format (1 µl per reaction) allowed the reaction chemistry to remain the same whilst the reaction volume, and therefore the required DNA input, to be greatly reduced i.e. the TaqMan Array assays running efficiently with only 50 ng DNA per 48 well channel (therefore approximately 1 ng DNA per reaction). This reduces the chance of assay failure due to too little DNA, while increasing the coverage of actionable mutations.

The second factor was the far superior discrimination/ identification power of the TaqMan Array plates. As each well on the plate only contains one assay i.e. the primers and probes

for one target, therefore the assay results are a very clear present or not present outcome (based on the presence or absence of a signal below our established threshold). Because of this there was a much lower occurrence of false positive or false negative results. This characteristic also means that interpretation of the results is far simpler than the equivalent process with the GeneFirst assays. The GeneFirst assays employed a very complicated system of ΔCt thresholds which were different depending on the targets and the master mixes. In particular, the GeneFirst KRAS version 5 assay introduced a new flow diagram for interpreting the ΔCt results from the two master mixes which was extremely difficult to interpret (see Methods sections 2.2.8 and figure 2.2.8.2). This type of analysis would be completely unsuitable for routine diagnostic use as the flow diagram was so ambiguous different operators may come to different interpretations of the assay results. By contrast the analysis of the TaqMan Array assay results were simple and could also be easily automated, removing the potential for operator error in the analysis step.

The final factor contributing to the TaqMan Array's superiority over the GeneFirst system is the user friendly enclosed 384 well plate format. With the plate separated into eight separate channels (48 wells per channel) the entire plate could be loaded in only eight pipetting actions maximum. This dramatically reduces hands on time, and most importantly greatly reduces the chances of accidental contamination by the operator or other operator errors. These three factors together resulted in excellent performance by the TaqMan Array with high sensitivity, specificity and reproducibility.

In addition to testing a large series of FFPE DNA samples from lung, colorectal and melanoma patients, the TA was also tested prospectively using a series of cfDNA samples from patient plasma. The results from this testing gave strong results (Ct<35) for all the TA control assays, which indicated the castPCR assays within the TA plate were robust and had a high tolerance

for potentially highly fragmented and low concentration DNA. Unfortunately there were no mutant targets that gave a signal below our threshold for positive results (Ct<36.5). There were two signals detected for *EGFR* and *NRAS*, however they were Ct37.9 and Ct36.9 and thus above the positive cut off set during validation. However examination of the amplification plots revealed the signals possessed genuine sigmoidal amplification profiles and thus there may be a case for adjusting the analysis thresholds for cfDNA samples.

5.4 CELL FREE DNA

The third approach employed in this project was to investigate if analysis of cfDNA could be practically and efficiently adopted by a diagnostic laboratory, by looking at the stability of cfDNA whilst stored in EDTA, by comparing extraction systems for isolating cfDNA, and finally to investigate if the cfDNA obtained in the routine laboratory could be used to detect oncogenic mutations with the Ion Torrent PGM platform.

The results from the evaluation of the extraction platforms was very helpful in terms of the very significant influence of extraction platform (and selection of an appropriate kit chemistry) on results. In a vast majority of cases throughout this evaluation, DNA extracts from the Promega Maxwell instrument produced superior results than those extracted by the Qiagen EZ1 platform. At initially DNA quantification, the Promega Maxwell extracted samples had consistently higher DNA concentrations, as measured by Qubit. When the samples were analysed by Ion Torrent NGS, the results generated by the Maxwell extracts were consistently superior in terms of the NGS analytics (read length, number of reads,

higher quality scores, higher coverage and more even strand bias). This is clear evidence that the DNA eluted from Maxwell was consistently of higher quality than that eluted by the EZ1.

In the broader context, this project shows that the quality of starting material is critical to successful of downstream molecular techniques, which is an important consideration when analysing the results from novel sources of DNA. This also highlights the point that if cfDNA were to be used diagnostically, the storage and separation procedures would be have to be as equally stringent as the measures taken when running the diagnostic assays, as otherwise the results generated may not accurately reflect the biological reality.

The issues related to processing methods of cfDNA highlighted by this project are analogous to the issues around the processing and storage of FFPE tissue, which is still the primary source of genetic material for molecular diagnostic analysis. Formalin has a number of irreversible effects on treated tissue, it dramatically slows autolysis and putrefaction by cross linking and denaturing proteins, and by fragmentation, cross linking and methylation of DNA (Hamazaki *et al.*, 1993; Koshiba *et al.*, 1993). Of particular concern in terms of molecular analysis is the artefact that formalin causes cross linking between cytosine molecules in the DNA, which in turn can cause DNA polymerases to fail to recognise the cytosine base and instead incorporate an adenosine or guanosine base, creating artificial mutations (Williams *et al.*, 1999). A number of studies of performed investigations into the effect of formalin fixation on downstream applications (Ferrer *et al.*, 2007; Miething *et al.*, 2006; Zsikla *et al.*, 2004) which universally conclude that formalin fixation does have an effect on molecular assays and that fixation procedures must be optimised to the specific sample type in order to balance preservation with good quality DNA for analysis. The issue of tissue fixation shows that no biological DNA source is ideal and that processing methods must be well optimised and efficiently performed to achieve the most accurate results in downstream tests.

Another important finding from this investigation was the importance of the choice of quantification method for DNA concentration in the DNA extracts. Spectrophotometry (used in the Nanodrop) is routinely used for DNA quantification in many diagnostic and routine settings, however the data from our investigation has showed this method to be very vulnerable to giving falsely high readings, most likely caused by samples with a high degree of contaminants (Kapp *et al.*, 2015). This suggests that in laboratories where Nanodrop is the primary or only method used to quantify DNA before molecular analysis, then there is a high likelihood that samples are not being loaded at the optimal concentration, leading to increased failure rates. Whilst the DNA may be sufficient to pass control assays, the low concentration of target template i.e. mutated tumour DNA, may be too low to generate a positive signal, resulting in a false negative result. The data from this section of the study supports the idea that molecular laboratories should use fluorescence based quantification methods (e.g. Qubit) rather than spectrophotometry based methods.

The most significant finding from the cfDNA investigation was that an actionable oncogenic mutation was identified from cfDNA using both the NGS platform and the TaqMan Array. The mutation was a *KRAS* p.G12C substitution which is a well-known hotspot mutation in colorectal cancer, and to a lesser extent in lung cancer. This finding shows that cfDNA does have the potential to be used to detect oncogenic mutations in clinical plasma samples, by not only NGS but also by TaqMan Array.

One clear finding from this investigation is that NGS is absolutely superior to PCR in terms of the depth of analysis, and can also provide more information about individual variants or mutations identified e.g. the frequency at which the mutant was detected as a proportion of all the reads from that amplicon. The disadvantage of NGS is that it is expensive and time

consuming, and thus is better suited to smaller throughput runs such as specialist testing. PCR on the other hand is rapid and relatively cheap. The disadvantage of PCR is that the number of targets it can detect is limited to the exact targets the test is designed for, and there is a relatively low upper limit (compared to NGS) of how many targets can be efficiently detected per run. This suggests that the ideal system in a pathology laboratory would be to test a majority of samples via a PCR array method, then use NGS to confirm a small number of samples that are of insufficient quality/ concentration for PCR analysis, or samples in which a driver mutation is not identified.

5.5 OUTCOMES FROM WORK PERFORMED IN THIS PROJECT

The major outcome from this project is the adoption of the validated TaqMan Array assay as the new diagnostic technique for the detection of activating *EGFR*, *KRAS*, *NRAS* and *BRAF* mutations in lung, colorectal and melanoma samples in the UHCW Pathology department, replacing the previously used Qiagen Therascreen assay (for *EGFR*) and the in-house castPCR assay (for *KRAS*). The new TA assay has combined both assays into one test, whilst adding additional capabilities in terms of detecting *NRAS* mutations and more *BRAF* mutations.

In addition to tissue, cfDNA is being trialled on the TaqMan Array as a prospective study (initial results to date shown in chapter four). Whilst this is in the early stages, in time this study will give an interesting insight into the potential for cfDNA to be used in TaqMan Array format assay. If this is successful then it may become possible to utilise cfDNA with PCR based methods rather than the more complex NGS techniques.

Beyond routine service, the validated TaqMan Array has been used for development of other diagnostic tools in the Pathology Department. The TaqMan Array (alongside NGS) was used as a comparator method for *BRAF* mutation detection against Qiagen Therascreen® and Roche Cobas® in order to validate a melanoma NGS panel. The study was subsequently published (Reiman *et al.*, 2017).

The final outcome from the project is the intention for UHCW Pathology to expand its services in order to include NGS analysis. A business case for this development, and to employ a new member of staff who will oversee/ perform this work has been accepted. This represents a considerable investment in terms of time and resources by UHCW Pathology, however adoption of NGS technology will dramatically update their molecular repertoire and significantly expand their molecular capabilities not just in terms of oncology but also potentially in microbiology, genetics and prenatal screening.

5.6 LIMTATIONS OF THE PROJECT

Whilst there were a number of positive outcomes from this investigation, there were also a number of limitations that had an impact on its overall success.

5.6.1 GeneFirst Assays

During the testing of the prototype GeneFirst PNA PCR assays, there was no guidance or feedback from the company regarding what aspects had been changed in the assays between each version of the assay. Recommended DNA inputs were guides rather than

optimised parameters, and this may have contributed to some of the assays failures or false results that were experienced during this phase of the project. Considering the highly variable nature of the assay performance, it may have been better to have used DNA extracted from cell lines to test these assays, rather than surplus clinical material. Although using surplus clinical FFPE samples was appropriate to the purpose the assays were designed for, in retrospect these samples would have been better used at a later stage when the assays were more reliable. Cell lines would also have been useful as this would allow the same lines to be re-tested with multiple versions of the assays, and therefore giving another means of analysing assay consistency. However unfortunately this was not a practical approach for the diagnostic laboratory due to limited culture facilities, and the need to keep human cell lines separate from microbial cultures. Using the surplus FFPE samples typically resulted in each sample only being analysed once, whereas being able to observe assay consistency over multiple runs would have been beneficial.

5.6.2 TaqMan Array

By contrast, there were few limitations to the validation of the TaqMan Array assay. The key limitation was the number of samples available to create the validation panels. Under ideal circumstances there would have been a higher number of lung, colorectal and melanoma samples tested during the validation. Thankfully there were enough to demonstrate the sensitivity and specificity of the assay and to complete the validation. Another issue with the validation was cost of the plates and the manufacturing time. As the plates are a custom layout, they need to be ordered and manufactured for the user each time, resulting in a delivery time of approximately six weeks for every order. In the event of any incident relating to optimal storage of the plates e.g. power failure to refrigeration equipment, a six week

replacement time would mean result in a considerable disruption to the service. To mitigate the possibility of service disruption and alternative method (Biocartis Idylla™) has now been brought into ensure continuity of service.

The final limitation of TA assay is limited by number of samples that can be run per plate. Only seven can be analysed at once. This is similar to the preceding assays (eight samples per run for Qiagen Therascreen) and at present suits the workload of the molecular laboratory, if demand for these services increases in the future then only being able to analyse seven samples at once may have an impact on turnaround time, although this is partially negated by the speed of the assay set up, meaning multiple runs could be performed per day with relative ease.

5.6.3 Plasma Samples

The greatest limitation in this entire project was undoubtedly the very limited access to patient plasma samples, which unfortunately was entirely out of the research group's control. In total over the course of three years only 40 samples were obtained, which represents an extremely small proportion of all the potential patients attending the UHCW lung clinic over the same period. There are several reasons for this. Firstly, limited support from clinical staff for the research work resulted in all the plasma samples coming from only two sources (the lung nursing team and one medic). Secondly, an inconsistent level of support from staff in the UHCW Pathology Blood Sciences laboratory meant on the rare occasions samples were sent up to our group, I was often not informed by Blood Sciences that the sample(s) had arrived, if I was informed at all. As a result, the number of plasma samples which were separated on day of receipt i.e. separated in 24 hours or less, was very

low. Many of the plasma samples which were separate from EDTA whole blood at 48 or 72 hours were due to us not being informed about the arrival of samples. So these two issues (i.e. low support from clinical staff and lack of support from Blood Sciences) exacerbated each other, resulting in low numbers of plasma samples received and sub-optimal storage of those that were received. A number of samples had to be discarded as they had been stored at room temperature for extended periods and haemolysed.

Another limitation with the cfDNA analysis was that no matched patient tissue samples were available for any of the plasma samples. Whilst this may be consistent with the frequent problem in lung cancer that tissue is difficult to obtain (especially at early stage), it would be have very interesting/ informative to have been able to sequence matching tissue and plasma samples from the same patients, as done in a number of other studies (Xu *et al.*, 2012).

5.7 DEVELOPMENTS IN THIS FIELD

Since the commencement of this investigation in 2013, the field of cfDNA analysis has grown dramatically. What was initially a novel approach is now on the verge of being used clinically. An online search of literature using 'cfDNA' as a key word yields 29 articles from January 2017 alone. Examination of these papers gives an interesting insight into the breadth of work being performed in this field. Amongst the 29 papers identified 20 of them are related to cancer, illustrating that cancer diagnostics are a key focus of cfDNA research. Within the 20 cancer papers, the articles were distributed amongst the following themes: lung (Cargnin *et al.*, 2017; Chen *et al.*, 2017; Lysov *et al.*, 2017; Malapelle *et al.*, 2017; Wu *et al.*, 2017); breast

(Page *et al.*, 2017; Shaw *et al.*, 2017); colorectal (Nagai *et al.*, 2017; Spindler, 2017), pancreatic (Allenson *et al.*, 2017; Woo *et al.*, 2017); other cancers (Kumari *et al.*, 2017; Rossi *et al.*, 2017; Sandulache *et al.*, 2017); general cancer or technical investigations (Fujii *et al.*, 2017; Kockan *et al.*, 2017; Ma *et al.*, 2017; Sorber *et al.*, 2017; Wang *et al.*, 2017); and one article on veterinary cancer (Beffagna *et al.*, 2017).

The articles of greatest relevance to this project were those related to lung and colorectal cancer. Of the five articles related to lung cancer, three focused on *EGFR* mutation detection with cfDNA utilising either RT-PCR (Malapelle *et al.*, 2017; Wu *et al.*, 2017) or NGS (Chen *et al.*, 2017). Malapelle et al (Malapelle *et al.*, 2017) evaluated the use of the Roche COBAS® EGFR mutation test v2 using cfDNA. They concluding that the test was a reliable and accurate test for use with both tissue and cfDNA. Wu et al (Wu *et al.*, 2017) analysed the serum, plasma and paired tissue samples from *EGFR* positive adenocarcinoma patients who were participating in a phase III clinical trial. The Qiagen Therascreen EGFR assay was used for mutation detection. The study reported that mutations were detected in 28.6% of serum and 60.5% of plasma, which supports the findings of other studies that favour plasma for cfDNA analysis. Chen et al (Chen *et al.*, 2017) used cfDNA to detect rare resistance mutations to osimertinib. CfDNA was isolated from plasma and from pleural effusion, and analysed using NGS. Comparison of mutation results from pre-treatment tissue revealed that resistance mutations were acquired after treatment. This article demonstrates the usefulness of cfDNA not only for diagnostics but for disease monitoring during treatment.

The other two articles related to lung cancer focused more of properties of cfDNA rather than its use for mutation detection. Cargnin et al (Cargnin *et al.*, 2017) performed a systematic review and meta-analysis of the impact of baseline cfDNA levels on survival outcomes in lung cancer patients. Whilst there was no statistically significant association

between baseline cfDNA level and progression free survival, the analysis of overall survival did reveal an increased risk of death with higher baseline cfDNA levels p <0.001 (Cargnin *et al.*, 2017). The study by Lysov et al (Lysov *et al.*, 2017) investigated the effect of platinum based chemotherapy on cfDNA levels, using non-tumour bearing murine control and a murine xenograph model. Comparisons of cfDNA levels in response to chemotherapy revealed that chemotherapy increased cfDNA in both healthy and in tumour-bearing mice.

The two papers related to colorectal cancer were a review by Spindler et al (Spindler, 2017) and a quantitative study of LINE-1 hypomethylation in cfDNA by Nagai et al (Nagai *et al.*, 2017). The review by Spindler et al concluded that studies using cfDNA in CRC are very heterogeneous but cfDNA is a very promising tool for diagnosis and prognosis of CRC. In the study by Nagai et al the group used cfDNA to determine the level of LINE-1 hypomethylation in 114 CRC patients. Their results showed that later stage patients had significantly higher LINE-1 hypomethylation levels compared to early stage, and that LINE-1 hypomethylation is higher in CRC patients compared to health controls (Nagai *et al.*, 2017). This article suggests LINE-1 hypomethylation (detected by cfDNA) could be a new biomarker for CRC.

Beyond lung and colorectal cancer there were 12 other publications related to cancer: two on breast cancer; two on pancreatic cancer; three on other cancers (gall bladder; B-cell lymphoma and anaplastic thyroid carcinoma) and five additional papers that did not focus on one type of cancer specifically. The two articles on breast cancer focused on investigating breast cancer genetic heterogeneity by NGS (Page *et al.*, 2017) and on comparing the genetic profiles of CTCs and cfDNA (Shaw *et al.*, 2017). The two articles on pancreatic cancer investigated exosome-derived DNA (Allenson *et al.*, 2017) and on the effect of chemotherapy on cfDNA in terms of concentration and mutations detectable (Woo *et al.*, 2017). The publication on gall bladder cancer quantified serum cfDNA levels in patients as potential

diagnostic marker (Kumari *et al.*, 2017). The study on B-cell lymphoma used NGS to genotype diffuse large B-cell lymphoma (DLBCL) from plasma cfDNA (Rossi *et al.*, 2017). Finally the paper on anaplastic thyroid carcinoma (ATC) compared the performance of two NGS assays for the genetic analysis of cfDNA from ATC patients (Sandulache *et al.*, 2017).

Outside the field of cancer, there were six articles on trisomy detection and three articles on other non-cancer cfDNA topics including: trauma care (Gogenur *et al.*, 2017); hepatic inflammation (Patrick *et al.*, 2017); and interstitial lung disease in dermatomyositis patients (Zhang *et al.*, 2017). Of the six articles related to trisomy detection, four were original studies of various aspects and methods for trisomy detection using cfDNA (Colosi *et al.*, 2017; Gerundino *et al.*, 2017; Hu *et al.*, 2017; Palomaki *et al.*, 2017), one was a review and meta-analysis of best practice for screening (Iwarsson *et al.*, 2017), and one article was the recommendations from the Society for Maternal-Fetal Medicine on use of ultra-sound with cfDNA testing (Norton *et al.*, 2017).

This brief summary of literature published in the first month of 2017 shows very clearly that there is a very high level of interest in the clinical utility of cfDNA, largely in cancer but followed closely by maternal-fetal medicine. The range of articles published on this subject also show that cfDNA has a very broad spectrum of applications, far beyond just cancer.

5.8 FUTURE WORK AND CONCLUSION

The relatively new field of Liquid Biopsy is an expanding area and interest in this field continues to grow. As a result there are a number of directions this work could follow.

In terms of technical approaches, there a number of ways the laboratory processes may be developed to expand or enhance the current capabilities. With reference to the TA assay, it has demonstrated robustness and efficiency in terms of the strong signals generated by the control assays when analysing cfDNA. The mutant assays did not perform as well, so one approach could involve pre-amplification of the template DNA to enrich the sample for the tumour DNA component. Whilst this may enhance the tumour derived DNA signal, there is a risk that pre-amplification could introduce non-specific products or amplification errors into the template DNA, so development of any technique such as this will have to be stringently monitored and quality assured to ensure test accuracy is maintained.

Activating and resistance to treatment mutations make up a majority of oncogenic changes in *EGFR*, however they are not the only considerations. *EGFR* amplification is another frequent occurrence in lung (Liang *et al.*, 2010) and colorectal cancer (Shen *et al.*, 2014), and detection of this phenomena should also be developed, both in terms of detection from tissue but also from plasma cfDNA.

Beyond *EGFR*, *ALK* fusions are another key concern for lung cancer diagnostics (Hainsworth & Anthony Greco, 2016). There are currently NGS panels designed to detect this phenomena in tissue, including a version of the Ion AmpliSeq™ RNA Fusion Lung Cancer Research Panel which is currently in development by Thermo Fisher, which was tested in the UHCW Pathology laboratory. Research has been conducted into detecting ALK fusions from cfDNA with promising results (Cui *et al.*, 2017; Dietz *et al.*, 2016).

The primary aim of future work would be to continue to develop the current methods for optimal detection of oncogenic mutations in cfDNA. This would not only involve lab based

optimisation, but also promoting the work to clinical staff to obtain more samples on a more regular basis, and to establish a more efficient system of obtaining samples through Blood Sciences, which caused a significant impact during this project.

The next aim would be to perform cfDNA monitoring on patients currently being treated at UHCW. Plasma could be used to monitor cfDNA levels and the tumour mutation profile of a patient at key monitoring points e.g. at investigation, during treatment, before and after surgery, and during recovery (where appropriate). Expansion of the role of cfDNA in this manner would give extremely important insight into the molecular/ genetic dynamics occurring during cancer treatment, and may also guide patient management and treatment decisions to produce the most favourable outcome.

Beyond the clinic, cfDNA may be useful as a screening tool. Analysing a population of high risk individuals (e.g. for lung cancer: chosen by age and smoking history) for mutations in cfDNA may have a benefit in terms of early diagnosis, potentially improving the patient outcomes and maximising treatment effectiveness. There may also be a saving to the health services in terms of more effective and cheaper early treatment compared to expensive and relatively ineffective therapy of more advanced disease.

The work presented in this thesis has a range of potentially useful clinical applications and should be developed further. Despite the multiple limitations present, the effectiveness and benefits of both high throughput diagnostic PCR platforms (TaqMan Array) and analysis of cfDNA have been demonstrated, with key findings being published in a peer reviewed journal. The TaqMan Array assay has introduced as the primary diagnostic test for detection of *EGFR*, *KRAS*, *NRAS* and *BRAF* mutations in the UHCW Pathology department. Also a business case has been submitted by UHCW Pathology and accepted by the Trust for the

introduction of Next Generation Sequencing as a diagnostic service in the department. Both these changes in UHCW Pathology have occurred due to work performed during this investigation. In the future liquid biopsy, in conjunction with sensitive and accurate molecular techniques, may provide a superior alternative to tissue biopsy for management of cancer patients, resulting in earlier diagnosis, more effective treatment and improved outcomes for patients.

REFERENCES

Ahmadian, A., Ehn, M. & Hober, S. (2006) Pyrosequencing: history, biochemistry and future. *Clin Chim Acta*, 363 (1-2): 83-94.

Albino, A. P., Le Strange, R., Oliff, A. I., Furth, M. E. & Old, L. J. (1984) Transforming ras genes from human melanoma: a manifestation of tumour heterogeneity? *Nature*, 308 (5954): 69-72.

Allegrini, S., Antona, J., Mezzapelle, R., Miglio, U., Paganotti, A., Veggiani, C., Frattini, M., Monga, G., Balbo, P. & Boldorini, R. (2012) Epidermal growth factor receptor gene analysis with a highly sensitive molecular assay in routine cytologic specimens of lung adenocarcinoma. *Am J Clin Pathol*, 138 (3): 377-381.

Allenson, K., Castillo, J., San Lucas, F. A., Scelo, G., Kim, D. U., Bernard, V., Davis, G., Kumar, T., Katz, M., Overman, M. J., Foretova, L., Fabianova, E., Holcatova, I., Janout, V., Meric-Bernstam, F., Gascoyne, P., Wistuba, I., Varadhachary, G., Brennan, P., Hanash, S., Li, D., Maitra, A. & Alvarez, H. (2017) High Prevalence of Mutant KRAS in Circulating Exosome-derived DNA from Early Stage Pancreatic Cancer Patients. *Ann Oncol*,

Angulo, B., Conde, E., Suarez-Gauthier, A., Plaza, C., Martinez, R., Redondo, P., Izquierdo, E., Rubio-Viqueira, B., Paz-Ares, L., Hidalgo, M. & Lopez-Rios, F. (2012) A comparison of EGFR mutation testing methods in lung carcinoma: direct sequencing, real-time PCR and immunohistochemistry. *PLoS One*, 7 (8): e43842.

Anker, P. & Stroun, M. (2000) Circulating DNA in plasma or serum. *Medicina (B Aires)*, 60 (5 Pt 2): 699-702.

Aqil, M., Naqvi, A. R., Mallik, S., Bandyopadhyay, S., Maulik, U. & Jameel, S. (2014) The HIV Nef protein modulates cellular and exosomal miRNA profiles in human monocytic cells. *J Extracell Vesicles*, 3

Arteaga, C. (2003) Targeting HER1/EGFR: a molecular approach to cancer therapy. *Semin Oncol*, 30 (3 Suppl 7): 3-14.

Aung, K. L., Board, R. E., Ellison, G., Donald, E., Ward, T., Clack, G., Ranson, M., Hughes, A., Newman, W. & Dive, C. (2010) Current status and future potential of somatic mutation testing from circulating free DNA in patients with solid tumours. *Hugo J*, 4 (1-4): 11-21.

Beffagna, G., Sammarco, A., Bedin, C., Romualdi, C., Mainenti, M., Mollo, A., Cavicchioli, L., Ferro, S., Trez, D., De Maria, R., Nitti, D., Saccani, A., Campanella, M., Agostini, M. & Zappulli, V. (2017) Circulating Cell-Free DNA in Dogs with

Mammary Tumors: Short and Long Fragments and Integrity Index. *PLoS One*, 12 (1): e0169454.

Berninger, H. (2010) *ARMS PCR in Molecular Pathology- Mutation Testing Made Simple*. IN: Molecular Diagnostics 2010. Conference:Location.

Betz, B. L., Roh, M. H., Weigelin, H. C., Placido, J. B., Schmidt, L. A., Farmen, S., Arenberg, D. A., Kalemkerian, G. P. & Knoepp, S. M. (2011) The application of molecular diagnostic studies interrogating EGFR and KRAS mutations to stained cytologic smears of lung carcinoma. *Am J Clin Pathol*, 136 (4): 564-571.

Billah, S., Stewart, J., Staerkel, G., Chen, S., Gong, Y. & Guo, M. (2011) EGFR and KRAS mutations in lung carcinoma: molecular testing by using cytology specimens. *Cancer Cytopathol*, 119 (2): 111-117.

Board, R. E., Wardley, A. M., Dixon, J. M., Armstrong, A. C., Howell, S., Renshaw, L., Donald, E., Greystoke, A., Ranson, M., Hughes, A. & Dive, C. (2010) Detection of PIK3CA mutations in circulating free DNA in patients with breast cancer. *Breast Cancer Res Treat*, 120 (2): 461-467.

Bolton, L., Reiman, A., Lucas, K., Timms, J. & Cree, I. A. (2015) KRAS mutation analysis by PCR: a comparison of two methods. *PLoS One*, 10 (1): e0115672.

Borras, E., Jurado, I., Hernan, I., Gamundi, M. J., Dias, M., Marti, I., Mane, B., Arcusa, A., Agundez, J. A., Blanca, M. & Carballo, M. (2011) Clinical pharmacogenomic testing of KRAS, BRAF and EGFR mutations by high resolution melting analysis and ultra-deep pyrosequencing. *BMC Cancer*, 11 406.

Breitbach, S., Tug, S. & Simon, P. (2012) Circulating cell-free DNA: an up-coming molecular marker in exercise physiology. *Sports Med*, 42 (7): 565-586.

Buttitta, F., Felicioni, L., Del Grammastro, M., Filice, G., Di Lorito, A., Malatesta, S., Viola, P., Centi, I., D'Antuono, T., Zappacosta, R., Rosini, S., Cuccurullo, F. & Marchetti, A. (2013) Effective assessment of egfr mutation status in bronchoalveolar lavage and pleural fluids by next-generation sequencing. *Clin Cancer Res*, 19 (3): 691-698.

Caen, O., Nizard, P., Garrigou, S., Perez-Toralla, K., Zonta, E., Laurent-Puig, P. & Taly, V. (2015) [Digital PCR compartmentalization II. Contribution for the quantitative detection of circulating tumor DNA]. *Med Sci (Paris)*, 31 (2): 180-186.

Cargnin, S., Canonico, P. L., Genazzani, A. A. & Terrazzino, S. (2017) Quantitative Analysis of Circulating Cell-Free DNA for Correlation with Lung Cancer Survival: A Systematic Review and Meta-Analysis. *J Thorac Oncol*, 12 (1): 43-53.

Carlos, G., Anforth, R., Clements, A., Menzies, A. M., Carlino, M. S., Chou, S. & Fernandez-Penas, P. (2015) Cutaneous Toxic Effects of BRAF Inhibitors Alone and in Combination With MEK Inhibitors for Metastatic Melanoma. *JAMA Dermatol*, 151 (10): 1103-1109.

Chen, K., Zhou, F., Shen, W., Jiang, T., Wu, X., Tong, X., Shao, Y. W., Qin, S. & Zhou, C. (2017) Novel mutations on EGFR Leu792 potentially correlate to acquired resistance to osimertinib in advanced NSCLC. *J Thorac Oncol*,

Chen, Z., Feng, J., Buzin, C. H., Liu, Q., Weiss, L., Kernstine, K., Somlo, G. & Sommer, S. S. (2009) Analysis of cancer mutation signatures in blood by a novel ultra-sensitive assay: monitoring of therapy or recurrence in non-metastatic breast cancer. *PLoS One*, 4 (9): e7220.

Colosi, E., D'Ambrosio, V. & Periti, E. (2017) First trimester contingent screening for trisomies 21,18,13: is this model cost efficient and feasible in public health system? *J Matern Fetal Neonatal Med*, 1-6.

Cree, I. A. (2016a) Diagnostic RAS mutation analysis by polymerase chain reaction (PCR). *Biomol Detect Quantif*, 8 29-32.

Cree, I. A. (2016b) Progress and potential of RAS mutation detection for diagnostics and companion diagnostics. *Expert Rev Mol Diagn*, 1-6.

Cree, I. A., Deans, Z., Ligtenberg, M. J., Normanno, N., Edsjo, A., Rouleau, E., Sole, F., Thunnissen, E., Timens, W., Schuuring, E., Dequeker, E., Murray, S., Dietel, M., Groenen, P., Van Krieken, J. H., European Society of Pathology Task Force on Quality Assurance in Molecular, P. & Royal College of, P. (2014) Guidance for laboratories performing molecular pathology for cancer patients. *J Clin Pathol*, 67 (11): 923-931.

Cui, S., Zhang, W., Xiong, L., Pan, F., Niu, Y., Chu, T., Wang, H., Zhao, Y. & Jiang, L. (2017) Use of capture-based next-generation sequencing to detect ALK fusion in plasma cell-free DNA of patients with non-small-cell lung cancer. *Oncotarget*, 8 (2): 2771-2780.

Cutler, R. E., Jr., Stephens, R. M., Saracino, M. R. & Morrison, D. K. (1998) Autoregulation of the Raf-1 serine/threonine kinase. *Proc Natl Acad Sci U S A*, 95 (16): 9214-9219.

Dahse, R., Berndt, A., Dahse, A. K. & Kosmehl, H. (2008a) Two allele-specific PCR assays for screening epidermal growth factor receptor gene hotspot mutations in lung adenocarcinoma. *Mol Med Rep*, 1 (1): 45-50.

Dahse, R., Berndt, A. & Kosmehl, H. (2008b) PCR-based testing for therapy-related EGFR mutations in patients with non-small cell lung cancer. *Anticancer Res*, 28 (4B): 2265-2270.

Daum, G., Eisenmann-Tappe, I., Fries, H. W., Troppmair, J. & Rapp, U. R. (1994) The ins and outs of Raf kinases. *Trends Biochem Sci*, 19 (11): 474-480.

Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M. J., Bottomley, W., Davis, N., Dicks, E., Ewing, R., Floyd, Y., Gray, K., Hall, S., Hawes, R., Hughes, J., Kosmidou, V., Menzies, A., Mould, C., Parker, A., Stevens, C., Watt, S., Hooper, S., Wilson, R., Jayatilake, H., Gusterson, B. A., Cooper, C., Shipley, J., Hargrave, D., Pritchard-Jones, K., Maitland, N., Chenevix-Trench, G., Riggins, G. J., Bigner, D. D., Palmieri, G., Cossu, A., Flanagan, A., Nicholson, A., Ho, J. W., Leung, S. Y., Yuen, S. T., Weber, B. L., Seigler, H. F., Darrow, T. L., Paterson, H., Marais, R., Marshall, C. J., Wooster, R., Stratton, M. R. & Futreal, P. A. (2002) Mutations of the BRAF gene in human cancer. *Nature*, 417 (6892): 949-954.

Devitt, B., Liu, W., Salemi, R., Wolfe, R., Kelly, J., Tzen, C. Y., Dobrovic, A. & McArthur, G. (2011) Clinical outcome and pathological features associated with NRAS mutation in cutaneous melanoma. *Pigment Cell Melanoma Res*, 24 (4): 666-672.

Dietz, S., Schirmer, U., Merce, C., von Bubnoff, N., Dahl, E., Meister, M., Muley, T., Thomas, M. & Sultmann, H. (2016) Low Input Whole-Exome Sequencing to Determine the Representation of the Tumor Exome in Circulating DNA of Non-Small Cell Lung Cancer Patients. *PLoS One*, 11 (8): e0161012.

Do, H., Krypuy, M., Mitchell, P. L., Fox, S. B. & Dobrovic, A. (2008) High resolution melting analysis for rapid and sensitive EGFR and KRAS mutation detection in formalin fixed paraffin embedded biopsies. *BMC Cancer*, 8 142.

Dobbie, F., Hiscock, R., Leonardi-Bee, J., Murray, S., Shahab, L., Aveyard, P., Coleman, T., McEwen, A., McRobbie, H., Purves, R. & Bauld, L. (2015) Evaluating Long-term Outcomes of NHS Stop Smoking Services (ELONS): a prospective cohort study. *Health Technol Assess*, 19 (95): 1-156.

Druliner, B. R., Rashtak, S., Ruan, X., Bae, T., Vasmatzis, N., O'Brien, D., Johnson, R., Felmlee-Devine, D., Washechek-Aletto, J., Basu, N., Liu, H., Smyrk, T., Abyzov, A. & Boardman, L. A. (2016) Colorectal Cancer with Residual Polyp of Origin: A Model of Malignant Transformation. *Transl Oncol*, 9 (4): 280-286.

Dubsky, P., Filipits, M., Jakesz, R., Rudas, M., Singer, C. F., Greil, R., Dietze, O., Luisser, I., Klug, E., Sedivy, R., Bachner, M., Mayr, D., Schmidt, M., Gehrmann, M. C., Petry, C., Weber, K. E., Kronenwett, R., Brase, J. C., Gnant, M., Austrian, B. & Colorectal Cancer Study, G. (2013) EndoPredict improves the prognostic classification derived from common clinical guidelines in ER-positive, HER2-negative early breast cancer. *Ann Oncol*, 24 (3): 640-647.

Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S. M., Driver, D. A., Berg, R. H., Kim, S. K., Norden, B. & Nielsen, P. E. (1993) PNA hybridizes to

complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature*, 365 (6446): 566-568.

El Zoghbi, M. & Cummings, L. C. (2016) New era of colorectal cancer screening. *World J Gastrointest Endosc*, 8 (5): 252-258.

Ellison, G., Zhu, G., Moulis, A., Dearden, S., Speake, G. & McCormack, R. (2013) EGFR mutation testing in lung cancer: a review of available methods and their use for analysis of tumour tissue and cytology samples. *J Clin Pathol*, 66 (2): 79-89.

Ettinger, D. S., Akerley, W., Bepler, G., Blum, M. G., Chang, A., Cheney, R. T., Chirieac, L. R., D'Amico, T. A., Demmy, T. L., Ganti, A. K., Govindan, R., Grannis, F. W., Jr., Jahan, T., Jahanzeb, M., Johnson, D. H., Kessinger, A., Komaki, R., Kong, F. M., Kris, M. G., Krug, L. M., Le, Q. T., Lennes, I. T., Martins, R., O'Malley, J., Osarogiagbon, R. U., Otterson, G. A., Patel, J. D., Pisters, K. M., Reckamp, K., Riely, G. J., Rohren, E., Simon, G. R., Swanson, S. J., Wood, D. E., Yang, S. C. & Members, N. N.-S. C. L. C. P. (2010) Non-small cell lung cancer. *J Natl Compr Canc Netw*, 8 (7): 740-801.

Favoriti, P., Carbone, G., Greco, M., Pirozzi, F., Pirozzi, R. E. & Corcione, F. (2016) Worldwide burden of colorectal cancer: a review. *Updates Surg*, 68 (1): 7-11.

Fearon, E. R. & Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell*, 61 (5): 759-767.

Fedorenko, I. V., Gibney, G. T. & Smalley, K. S. (2013) NRAS mutant melanoma: biological behavior and future strategies for therapeutic management. *Oncogene*, 32 (25): 3009-3018.

Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C. & Parkin, D. M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*, 127 (12): 2893-2917.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D. & Bray, F. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*, 136 (5): E359-386.

Ferrer, I., Armstrong, J., Capellari, S., Parchi, P., Arzberger, T., Bell, J., Budka, H., Strobel, T., Giaccone, G., Rossi, G., Bogdanovic, N., Fakai, P., Schmitt, A., Riederers, P., Al-Sarraj, S., Ravid, R. & Kretzschmar, H. (2007) Effects of formalin fixation, paraffin embedding, and time of storage on DNA preservation in brain tissue: a BrainNet Europe study. *Brain Pathol*, 17 (3): 297-303.

Field, J. K., Duffy, S. W., Baldwin, D. R., Brain, K. E., Devaraj, A., Eisen, T., Green, B. A., Holemans, J. A., Kavanagh, T., Kerr, K. M., Ledson, M., Lifford, K. J., McRonald, F. E., Nair, A., Page, R. D., Parmar, M. K., Rintoul, R. C., Screaton, N.,

Wald, N. J., Weller, D., Whynes, D. K., Williamson, P. R., Yadegarfar, G. & Hansell, D. M. (2016) The UK Lung Cancer Screening Trial: a pilot randomised controlled trial of low-dose computed tomography screening for the early detection of lung cancer. *Health Technol Assess*, 20 (40): 1-146.

Forshew, T., Murtaza, M., Parkinson, C., Gale, D., Tsui, D. W., Kaper, F., Dawson, S. J., Piskorz, A. M., Jimenez-Linan, M., Bentley, D., Hadfield, J., May, A. P., Caldas, C., Brenton, J. D. & Rosenfeld, N. (2012) Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med*, 4 (136): 136ra168.

Fujii, T., Barzi, A., Sartore-Bianchi, A., Cassingena, A., Siravegna, G., Karp, D., Piha-Paul, S., Subbiah, V., Tsimberidou, A. M., Huang, H., Veronese, S., Di Nicolantonio, F., Pingle, S. C., Vibat, C. R., Hancock, S., Berz, D., Melnikova, V. O., Erlander, M. G., Luthra, R., Kopetz, S., Meric-Bernstam, F., Siena, S., Lenz, H. J., Bardelli, A. & Janku, F. (2017) Mutation-Enrichment Next-Generation Sequencing for Quantitative Detection of KRAS Mutations in Urine Cell-Free DNA from Patients with Advanced Cancers. *Clin Cancer Res*,

Fujita, S., Masago, K., Takeshita, J., Okuda, C., Otsuka, K., Hata, A., Kaji, R., Katakami, N. & Hirata, Y. (2015) Validation of an Ion Torrent Sequencing Platform for the Detection of Gene Mutations in Biopsy Specimens from Patients with Non-Small-Cell Lung Cancer. *PLoS One*, 10 (6): e0130219.

Galeazzi, M., Morozzi, G., Piccini, M., Chen, J., Bellisai, F., Fineschi, S. & Marcolongo, R. (2003) Dosage and characterization of circulating DNA: present usage and possible applications in systemic autoimmune disorders. *Autoimmun Rev*, 2 (1): 50-55.

Garrigou, S., Perkins, G., Garlan, F., Normand, C., Didelot, A., Le Corre, D., Peyvandi, S., Mulot, C., Niarra, R., Aucouturier, P., Chatellier, G., Nizard, P., Perez-Toralla, K., Zonta, E., Charpy, C., Pujals, A., Barau, C., Bouche, O., Emile, J. F., Pezet, D., Bibeau, F., Hutchison, J. B., Link, D. R., Zaanan, A., Laurent-Puig, P., Sobhani, I. & Taly, V. (2016) A Study of Hypermethylated Circulating Tumor DNA as a Universal Colorectal Cancer Biomarker. *Clin Chem*,

Gerundino, F., Giachini, C., Contini, E., Benelli, M., Marseglia, G., Giuliani, C., Marin, F., Nannetti, G., Lisi, E., Sbernini, F., Periti, E., Cordisco, A., Colosi, E., D'Ambrosio, V., Mazzi, M., Rossi, M., Staderini, L., Minuti, B., Pelo, E., Cicatiello, R., Maruotti, G. M., Sglavo, G., Conti, A., Frusconi, S., Pescucci, C. & Torricelli, F. (2017) Validation of a method for noninvasive prenatal testing for fetal aneuploidies risk and considerations for its introduction in the Public Health System. *J Matern Fetal Neonatal Med*, 30 (6): 710-716.

Ghorbian, S. & Ardekani, A. M. (2012) Non-Invasive Detection of Esophageal Cancer using Genetic Changes in Circulating Cell-Free DNA. *Avicenna J Med Biotechnol*, 4 (1): 3-13.

Glaysher, S., Yiannakis, D., Gabriel, F. G., Johnson, P., Polak, M. E., Knight, L. A., Goldthorpe, Z., Peregrin, K., Gyi, M., Modi, P., Rahamim, J., Smith, M. E., Amer, K., Addis, B., Poole, M., Narayanan, A., Gulliford, T. J., Andreotti, P. E. & Cree, I. A. (2009) Resistance gene expression determines the in vitro chemosensitivity of non-small cell lung cancer (NSCLC). *BMC Cancer*, 9 300.

Gogenur, M., Burcharth, J. & Gogenur, I. (2017) The role of total cell-free DNA in predicting outcomes among trauma patients in the intensive care unit: a systematic review. *Crit Care*, 21 (1): 14.

Gonzalez-Cao, M., Mayo-de-Las-Casas, C., Molina-Vila, M. A., De Mattos-Arruda, L., Munoz-Couselo, E., Manzano, J. L., Cortes, J., Berros, J. P., Drozdowskyj, A., Sanmamed, M., Gonzalez, A., Alvarez, C., Viteri, S., Karachaliou, N., Martin Algarra, S., Bertran-Alamillo, J., Jordana-Ariza, N. & Rosell, R. (2015) BRAF mutation analysis in circulating free tumor DNA of melanoma patients treated with BRAF inhibitors. *Melanoma Res*, 25 (6): 486-495.

Gormally, E., Caboux, E., Vineis, P. & Hainaut, P. (2007) Circulating free DNA in plasma or serum as biomarker of carcinogenesis: practical aspects and biological significance. *Mutat Res*, 635 (2-3): 105-117.

Grady, W. M. & Markowitz, S. D. (2015) The molecular pathogenesis of colorectal cancer and its potential application to colorectal cancer screening. *Dig Dis Sci*, 60 (3): 762-772.

Hagiwara, K. & Kobayashi, K. (2013) Importance of the cytological samples for the epidermal growth factor receptor gene mutation test for non-small cell lung cancer. *Cancer Sci*, 104 (3): 291-297.

Hainsworth, J. D. & Anthony Greco, F. (2016) Lung Adenocarcinoma with Anaplastic Lymphoma Kinase (ALK) Rearrangement Presenting as Carcinoma of Unknown Primary Site: Recognition and Treatment Implications. *Drugs Real World Outcomes*, 3 115-120.

Hamazaki, S., Koshiba, M., Habuchi, T., Takahashi, R. & Sugiyama, T. (1993) The effect of formalin fixation on restriction endonuclease digestion of DNA and PCR amplification. *Pathol Res Pract*, 189 (5): 553-557.

Higuchi, R., Fockler, C., Dollinger, G. & Watson, R. (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y)*, 11 (9): 1026-1030.

Hirsch, F. R., Varella-Garcia, M., Bunn, P. A., Jr., Di Maria, M. V., Veve, R., Bremmes, R. M., Baron, A. E., Zeng, C. & Franklin, W. A. (2003) Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J Clin Oncol*, 21 (20): 3798-3807.

Holland, P. M., Abramson, R. D., Watson, R. & Gelfand, D. H. (1991) Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci U S A*, 88 (16): 7276-7280.

HQO, H. Q. O. (2009) Fecal occult blood test for colorectal cancer screening: an evidence-based analysis. *Ont Health Technol Assess Ser*, 9 (10): 1-40.

Hu, C., Liu, X., Chen, Y., Sun, X., Gong, Y., Geng, M. & Bi, L. (2012) Direct serum and tissue assay for EGFR mutation in non-small cell lung cancer by high-resolution melting analysis. *Oncol Rep*, 28 (5): 1815-1821.

Hu, H., Jiang, Y., Zhang, M., Liu, S., Hao, N., Zhou, J., Liu, J., Zhang, X. & Ma, L. (2017) A prospective clinical trial to compare the performance of dried blood spots prenatal screening for Down's syndrome with conventional non-invasive testing technology. *Exp Biol Med (Maywood)*, 1535370216683837.

Hyun, K. A., Kim, J., Gwak, H. & Jung, H. I. (2016) Isolation and enrichment of circulating biomarkers for cancer screening, detection, and diagnostics. *Analyst*, 141 (2): 382-392.

Ishii, H., Azuma, K., Sakai, K., Kawahara, A., Yamada, K., Tokito, T., Okamoto, I., Nishio, K. & Hoshino, T. (2015) Digital PCR analysis of plasma cell-free DNA for non-invasive detection of drug resistance mechanisms in EGFR mutant NSCLC: Correlation with paired tumor samples. *Oncotarget*,

Iwarsson, E., Jacobsson, B., Dagerhamn, J., Davidson, T., Bernabe, E. & Heibert Arnlind, M. (2017) Analysis of cell-free fetal DNA in maternal blood for detection of trisomy 21, 18 and 13 in a general pregnant population and in a high risk population - a systematic review and meta-analysis. *Acta Obstet Gynecol Scand*, 96 (1): 7-18.

Jardines, L., Weiss, M., Fowble, B. & Greene, M. (1993) neu(c-erbB-2/HER2) and the epidermal growth factor receptor (EGFR) in breast cancer. *Pathobiology*, 61 (5-6): 268-282.

Jensen, K. K., Orum, H., Nielsen, P. E. & Norden, B. (1997) Kinetics for hybridization of peptide nucleic acids (PNA) with DNA and RNA studied with the BIAcore technique. *Biochemistry*, 36 (16): 5072-5077.

Jiang, N., Liu, N., Yang, F., Zhou, Q., Cui, R., Jiang, W., He, Q., Li, W., Guo, Y., Zeng, J., Yun, J., Chen, X., Zhou, B., Sun, Y., Wang, H., Chen, Z. G. & Ma, J. (2014) Hotspot mutations in common oncogenes are infrequent in nasopharyngeal carcinoma. *Oncol Rep*, 32 (4): 1661-1669.

Johnson, C. M., Wei, C., Ensor, J. E., Smolenski, D. J., Amos, C. I., Levin, B. & Berry, D. A. (2013) Meta-analyses of colorectal cancer risk factors. *Cancer Causes Control*, 24 (6): 1207-1222.

Johnson, P. J. & Lo, Y. M. (2002) Plasma nucleic acids in the diagnosis and management of malignant disease. *Clin Chem*, 48 (8): 1186-1193.

Joshi, S. S. & Arankalle, V. A. (2015) Enhanced humoral response in pregnant mice immunized with liposome encapsulated recombinant neutralizing epitope protein of Hepatitis- E virus. *Virol J*, 12 70.

Kalemkerian, G. P., Akerley, W., Bogner, P., Borghaei, H., Chow, L. Q., Downey, R. J., Gandhi, L., Ganti, A. K., Govindan, R., Grecula, J. C., Hayman, J., Heist, R. S., Horn, L., Jahan, T., Koczywas, M., Loo, B. W., Jr., Merritt, R. E., Moran, C. A., Niell, H. B., O'Malley, J., Patel, J. D., Ready, N., Rudin, C. M., Williams, C. C., Jr., Gregory, K., Hughes, M. & National Comprehensive Cancer, N. (2013) Small cell lung cancer. *J Natl Compr Canc Netw*, 11 (1): 78-98.

Kamel-Reid, S., Chong, G., Ionescu, D. N., Magliocco, A. M., Spatz, A., Tsao, M., Weng, X., Young, S., Zhang, T. & Soulieres, D. (2012) EGFR tyrosine kinase mutation testing in the treatment of non-small-cell lung cancer. *Curr Oncol*, 19 (2): e67-74.

Kapp, J. R., Diss, T., Spicer, J., Gandy, M., Schrijver, I., Jennings, L. J., Li, M. M., Tsongalis, G. J., de Castro, D. G., Bridge, J. A., Wallace, A., Deignan, J. L., Hing, S., Butler, R., Verghese, E., Latham, G. J. & Hamoudi, R. A. (2015) Variation in pre-PCR processing of FFPE samples leads to discrepancies in BRAF and EGFR mutation detection: a diagnostic RING trial. *J Clin Pathol*, 68 (2): 111-118.

Karachaliou, N., Mayo-de las Casas, C., Queralt, C., de Aguirre, I., Melloni, B., Cardenal, F., Garcia-Gomez, R., Massuti, B., Sanchez, J. M., Porta, R., Ponce-Aix, S., Moran, T., Carcereny, E., Felip, E., Bover, I., Insa, A., Reguart, N., Isla, D., Vergnenegre, A., de Marinis, F., Gervais, R., Corre, R., Paz-Ares, L., Morales-Espinosa, D., Viteri, S., Drozdowskyj, A., Jordana-Ariza, N., Ramirez-Serrano, J. L., Molina-Vila, M. A., Rosell, R. & Spanish Lung Cancer, G. (2015) Association of EGFR L858R Mutation in Circulating Free DNA With Survival in the EURTAC Trial. *JAMA Oncol*, 1 (2): 149-157.

Kikuchi H., R. A., Nyoni, J., Lloyd K., Savage R., Wotherspoon T., Berry L., Snead D., Cree I.A. (2016) Development and validation of a Taqman array for cancer mutation analysis. *Pathogenesis*, In press

Kim, H. R., Lee, S. Y., Hyun, D. S., Lee, M. K., Lee, H. K., Choi, C. M., Yang, S. H., Kim, Y. C., Lee, Y. C., Kim, S. Y., Jang, S. H., Lee, J. C. & Lee, K. Y. (2013) Detection of EGFR mutations in circulating free DNA by PNA-mediated PCR clamping. *J Exp Clin Cancer Res*, 32 (1): 50.

Kimura, H., Fujiwara, Y., Sone, T., Kunitoh, H., Tamura, T., Kasahara, K. & Nishio, K. (2006) EGFR mutation status in tumour-derived DNA from pleural effusion fluid is a practical basis for predicting the response to gefitinib. *Br J Cancer*, 95 (10): 1390-1395.

Kockan, C., Hach, F., Sarrafi, I., Bell, R. H., McConeghy, B., Beja, K., Haegert, A., Wyatt, A. W., Volik, S. V., Chi, K. N., Collins, C. C. & Sahinalp, S. C. (2017) SiNVICT: ultra-sensitive detection of single nucleotide variants and indels in circulating tumour DNA. *Bioinformatics*, 33 (1): 26-34.

Koshiba, M., Ogawa, K., Hamazaki, S., Sugiyama, T., Ogawa, O. & Kitajima, T. (1993) The effect of formalin fixation on DNA and the extraction of high-molecular-weight DNA from fixed and embedded tissues. *Pathol Res Pract*, 189 (1): 66-72.

Koudelakova, V., Kneblova, M., Trojanec, R., Drabek, J. & Hajduch, M. (2013) Non-small cell lung cancer - genetic predictors. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*, 157 (2): 125-136.

Kumar, R., Singh, S. K., Koshkin, A. A., Rajwanshi, V. K., Meldgaard, M., Wengel, J. (1998) The first analogues of LNA (locked nucleic acids): phosphorothioate-LNA and 2'-thio-LNA. *Bioorg Med Chem Lett.* 18;8 (16): 2219-22.

Kumari, S., Tewari, S., Husain, N., Agarwal, A., Pandey, A., Singhal, A. & Lohani, M. (2017) Quantification of Circulating Free DNA as a Diagnostic Marker in Gall Bladder Cancer. *Pathol Oncol Res*, 23 (1): 91-97.

Lee, C., Kang, K. H., Koh, Y., Chang, J., Chung, H. S., Park, S. K., Yoo, K. & Song, J. S. (2000) Characteristics of lung cancer in Korea, 1997. *Lung Cancer*, 30 (1): 15-22.

Lee, S. C., Lim, S. G., Soo, R., Hsieh, W. S., Guo, J. Y., Putti, T., Tao, Q., Soong, R. & Goh, B. C. (2006) Lack of somatic mutations in EGFR tyrosine kinase domain in hepatocellular and nasopharyngeal carcinoma. *Pharmacogenet Genomics*, 16 (1): 73-74.

Lee, S. M. & Chow, L. Q. (2014) A new addition to the PD-1 checkpoint inhibitors for non-small cell lung cancer-the anti-PDL1 antibody-MEDI4736. *Transl Lung Cancer Res*, 3 (6): 408-410.

Lee, T. H., Montalvo, L., Chrebtow, V. & Busch, M. P. (2001) Quantitation of genomic DNA in plasma and serum samples: higher concentrations of genomic DNA found in serum than in plasma. *Transfusion*, 41 (2): 276-282.

Leon, S. A., Shapiro, B., Sklaroff, D. M. & Yaros, M. J. (1977) Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res*, 37 (3): 646-650.

Leslie, M. (2011) Mysteries of the cell. How does a cell know its size? *Science*, 334 (6059): 1047-1048.

Li, W., Han, M. & Guan, K. L. (2000) The leucine-rich repeat protein SUR-8 enhances MAP kinase activation and forms a complex with Ras and Raf. *Genes Dev*, 14 (8): 895-900.

Liang, Z., Zhang, J., Zeng, X., Gao, J., Wu, S. & Liu, T. (2010) Relationship between EGFR expression, copy number and mutation in lung adenocarcinomas. *BMC Cancer*, 10 376.

Liu, X., Lu, Y., Zhu, G., Lei, Y., Zheng, L., Qin, H., Tang, C., Ellison, G., McCormack, R. & Ji, Q. (2013a) The diagnostic accuracy of pleural effusion and plasma samples versus tumour tissue for detection of EGFR mutation in patients with advanced non-small cell lung cancer: comparison of methodologies. *J Clin Pathol*, 66 (12): 1065-1069.

Liu, X., Lu, Y., Zhu, G., Lei, Y., Zheng, L., Qin, H., Tang, C., Ellison, G., McCormack, R. & Ji, Q. (2013b) The diagnostic accuracy of pleural effusion and plasma samples versus tumour tissue for detection of EGFR mutation in patients with advanced non-small cell lung cancer: comparison of methodologies. *J Clin Pathol*,

Lo, Y. M., Zhang, J., Leung, T. N., Lau, T. K., Chang, A. M. & Hjelm, N. M. (1999) Rapid clearance of fetal DNA from maternal plasma. *Am J Hum Genet*, 64 (1): 218-224.

Lopez-Rios, F., Angulo, B., Gomez, B., Mair, D., Martinez, R., Conde, E., Shieh, F., Tsai, J., Vaks, J., Current, R., Lawrence, H. J. & Gonzalez de Castro, D. (2013) Comparison of molecular testing methods for the detection of EGFR mutations in formalin-fixed paraffin-embedded tissue specimens of non-small cell lung cancer. *J Clin Pathol*, 66 (5): 381-385.

Lysov, Z., Dwivedi, D. J., Gould, T. J. & Liaw, P. C. (2017) Procoagulant effects of lung cancer chemotherapy: impact on microparticles and cell-free DNA. *Blood Coagul Fibrinolysis*, 28 (1): 72-82.

Ma, X., Zhu, L., Wu, X., Bao, H., Wang, X., Chang, Z., Shao, Y. W. & Wang, Z. (2017) Cell-Free DNA Provides a Good Representation of the Tumor Genome Despite Its Biased Fragmentation Patterns. *PLoS One*, 12 (1): e0169231.

Malapelle, U., Bellevicine, C., De Luca, C., Salatiello, M., De Stefano, A., Rocco, D., de Rosa, N., Vitiello, F., Russo, S., Pepe, F., Iaccarino, A., Micheli, P., Illiano, A., Carlomagno, C., Piantedosi, F. V. & Troncone, G. (2013) EGFR mutations detected on cytology samples by a centralized laboratory reliably predict response to gefitinib in non-small cell lung carcinoma patients. *Cancer Cytopathol*, 121 (10): 552-560.

Malapelle, U., de Luca, C., Vigliar, E., Ambrosio, F., Rocco, D., Pisapia, P., Bellevicine, C. & Troncone, G. (2016) EGFR mutation detection on routine cytological smears of non-small cell lung cancer by digital PCR: a validation study. *J Clin Pathol*, 69 (5): 454-457.

Malapelle, U., de Rosa, N., Rocco, D., Bellevicine, C., Crispino, C., Illiano, A., Piantedosi, F. V., Nappi, O. & Troncone, G. (2012) EGFR and KRAS mutations detection on lung cancer liquid-based cytology: a pilot study. *J Clin Pathol*, 65 (1): 87-91.

Malapelle, U., Sirera, R., Jantus-Lewintre, E., Reclusa, P., Calabuig-Farinas, S., Blasco, A., Pisapia, P., Rolfo, C. & Camps, C. (2017) Profile of the Roche cobas(R) EGFR mutation test v2 for non-small cell lung cancer. *Expert Rev Mol Diagn*,

Mandel, P. & Metais, P. (1948) Les acides nucleiques du plasma sanguin chez l'homme. *C R Seances Soc Biol Fil*, 142 (3-4): 241-243.

Marshall, C. J., Hall, A. & Weiss, R. A. (1982) A transforming gene present in human sarcoma cell lines. *Nature*, 299 (5879): 171-173.

Martins, G. A., Kawamura, M. T. & Carvalho Mda, G. (2000) Detection of DNA in the plasma of septic patients. *Ann N Y Acad Sci*, 906 134-140.

McGrath, J. P., Capon, D. J., Smith, D. H., Chen, E. Y., Seeburg, P. H., Goeddel, D. V. & Levinson, A. D. (1983) Structure and organization of the human Ki-ras proto-oncogene and a related processed pseudogene. *Nature*, 304 (5926): 501-506.

Mead, R., Duku, M., Bhandari, P. & Cree, I. A. (2011) Circulating tumour markers can define patients with normal colons, benign polyps, and cancers. *Br J Cancer*, 105 (2): 239-245.

Miething, F., Hering, S., Hanschke, B. & Dressler, J. (2006) Effect of fixation to the degradation of nuclear and mitochondrial DNA in different tissues. *J Histochem Cytochem*, 54 (3): 371-374.

Milagre, C., Dhomen, N., Geyer, F. C., Hayward, R., Lambros, M., Reis-Filho, J. S. & Marais, R. (2010) A mouse model of melanoma driven by oncogenic KRAS. *Cancer Res*, 70 (13): 5549-5557.

Mitri, Z., Constantine, T. & O'Regan, R. (2012) The HER2 Receptor in Breast Cancer: Pathophysiology, Clinical Use, and New Advances in Therapy. *Chemother Res Pract*, 2012 743193.

Mok, T., Wu, Y. L., Lee, J. S., Yu, C. J., Sriuranpong, V., Sandoval-Tan, J., Ladrera, G., Thongprasert, S., Srimuninnimit, V., Liao, M., Zhu, Y., Zhou, C., Fuerte, F., Margono, B., Wen, W., Tsai, J., Truman, M., Klughammer, B., Shames, D. S. & Wu,

L. (2015) Detection and Dynamic Changes of EGFR Mutations from Circulating Tumor DNA as a Predictor of Survival Outcomes in NSCLC Patients Treated with First-line Intercalated Erlotinib and Chemotherapy. *Clin Cancer Res*, 21 (14): 3196-3203.

Moore, H. M., Kelly, A. B., Jewell, S. D., McShane, L. M., Clark, D. P., Greenspan, R., Hayes, D. F., Hainaut, P., Kim, P., Mansfield, E. A., Potapova, O., Riegman, P., Rubinstein, Y., Seijo, E., Somiari, S., Watson, P., Weier, H. U., Zhu, C. & Vaught, J. (2011) Biospecimen reporting for improved study quality (BRISQ). *Cancer Cytopathol*, 119 (2): 92-101.

Morgan, S. R., Whiteley, J., Donald, E., Smith, J., Eisenberg, M. T., Kallam, E. & Kam-Morgan, L. (2012) Comparison of KRAS Mutation Assessment in Tumor DNA and Circulating Free DNA in Plasma and Serum Samples. *Clin Med Insights Pathol*, 5 15-22.

Muhammad, S. A., Fatima, N., Syed, N. I., Wu, X., Yang, X. F. & Chen, J. Y. (2015) MicroRNA Expression Profiling of Human Respiratory Epithelium Affected by Invasive Candida Infection. *PLoS One*, 10 (8): e0136454.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*, 51 Pt 1 263-273.

Nagai, Y., Sunami, E., Yamamoto, Y., Hata, K., Okada, S., Murono, K., Yasuda, K., Otani, K., Nishikawa, T., Tanaka, T., Kiyomatsu, T., Kawai, K., Nozawa, H., Ishihara, S., Hoon, D. S. & Watanabe, T. (2017) LINE-1 hypomethylation status of circulating cell-free DNA in plasma as a biomarker for colorectal cancer. *Oncotarget*,

Nakamura, H., Kawasaki, N., Taguchi, M. & Kabasawa, K. (2006) Survival impact of epidermal growth factor receptor overexpression in patients with non-small cell lung cancer: a meta-analysis. *Thorax*, 61 (2): 140-145.

NCBI (2016) *BRAF B-Raf proto-oncogene, serine/threonine kinase [Homo sapiens (human)]*. [online] Available from: https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowDetailView&TermToSearch=673 (Accessed 03/10/2016).

NCBI (2017) *ERBB4 v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 4 [Homo sapiens (human)]* [online] Available from: https://www.ncbi.nlm.nih.gov/gene/2066 (Accessed 23/03).

NICE (2011a) *Colorectal cancer: diagnosis and management. Clinical guideline [CG131]*. [online] Available from: https://www.nice.org.uk/guidance/cg131?unlid=986044020201541754732 (Accessed 04/10).

NICE (2011b) National Institute for Health and Clinical Excellence, Clinical Guideline CG121 Lung cancer: The diagnosis and treatment of lung cancer.

NICE (2012) National Institute for Health and Clinical Excellence Diagnostics Assessment Programme: Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer. Final scope.

NICE (2015) KRAS mutation testing of tumours in adults with metastatic colorectal cancer (discontinued).

NICE (2016) Colorectal cancer (metastatic) - cetuximab (review TA176) and panitumumab (part review TA240) (1st line) ID794.

NICE (2017) *NICE Pathway: Diagnosis and Staging of Lung Cancer*. [online] Available from: https://pathways.nice.org.uk/pathways/lung-cancer#path=view%3A/pathways/lung-cancer/diagnosis-and-staging-of-lung-cancer.xml&content=view-index (Accessed 09/03).

Nielsen, P. E. & Egholm, M. (1999) An introduction to peptide nucleic acid. *Curr Issues Mol Biol*, 1 (1-2): 89-104.

Norton, M. E., Biggio, J. R., Kuller, J. A. & Blackwell, S. C. (2017) SMFM Consult Series #42: The role of ultrasound in women who undergo cell-free DNA screening. *Am J Obstet Gynecol*,

Page, K., Guttery, D. S., Fernandez-Garcia, D., Hills, A., Hastings, R. K., Luo, J., Goddard, K., Shahin, V., Woodley-Barker, L., Rosales, B. M., Coombes, R. C., Stebbing, J. & Shaw, J. A. (2017) Next Generation Sequencing of Circulating Cell-Free DNA for Evaluating Mutations and Gene Amplification in Metastatic Breast Cancer. *Clin Chem*, 63 (2): 532-541.

Palomaki, G. E., Kloza, E. M., O'Brien, B. M., Eklund, E. E. & Lambert-Messerlian, G. M. (2017) The clinical utility of DNA-based screening for fetal aneuploidy by primary obstetrical care providers in the general pregnancy population. *Genet Med*,

Pan, J., Kong, L., Lin, S., Chen, G., Chen, Q. & Lu, J. J. (2008) The clinical significance of coexpression of cyclooxygenases-2, vascular endothelial growth factors, and epidermal growth factor receptor in nasopharyngeal carcinoma. *Laryngoscope*, 118 (11): 1970-1975.

Pan, Q., Pao, W. & Ladanyi, M. (2005) Rapid polymerase chain reaction-based detection of epidermal growth factor receptor gene mutations in lung adenocarcinomas. *J Mol Diagn*, 7 (3): 396-403.

Park, J. L., Kim, H. J., Choi, B. Y., Lee, H. C., Jang, H. R., Song, K. S., Noh, S. M., Kim, S. Y., Han, D. S. & Kim, Y. S. (2012) Quantitative analysis of cell-free DNA in the plasma of gastric cancer patients. *Oncol Lett*, 3 (4): 921-926.

Pasay, C., Arlian, L., Morgan, M., Vyszenski-Moher, D., Rose, A., Holt, D., Walton, S. & McCarthy, J. (2008) High-resolution melt analysis for the detection of a mutation associated with permethrin resistance in a population of scabies mites. *Med Vet Entomol*, 22 (1): 82-88.

Patrick, A. L., Grin, P. M., Kraus, N., Gold, M., Berardocco, M., Liaw, P. C., Fox-Robichaud, A. E. & Canadian Critical Care Translational Biology, G. (2017) Resuscitation fluid composition affects hepatic inflammation in a murine model of early sepsis. *Intensive Care Med Exp*, 5 (1): 5.

Pichler, M., Balic, M., Stadelmeyer, E., Ausch, C., Wild, M., Guelly, C., Bauernhofer, T., Samonigg, H., Hoefler, G. & Dandachi, N. (2009) Evaluation of high-resolution melting analysis as a diagnostic tool to detect the BRAF V600E mutation in colorectal tumors. *J Mol Diagn*, 11 (2): 140-147.

Pikor, L. A., Ramnarine, V. R., Lam, S. & Lam, W. L. (2013) Genetic alterations defining NSCLC subtypes and their therapeutic implications. *Lung Cancer*, 82 (2): 179-189.

Prior, I. A. & Hancock, J. F. (2012) Ras trafficking, localization and compartmentalized signalling. *Semin Cell Dev Biol*, 23 (2): 145-153.

Pulciani, S., Santos, E., Lauver, A. V., Long, L. K., Robbins, K. C. & Barbacid, M. (1982) Oncogenes in human tumor cell lines: molecular cloning of a transforming gene from human bladder carcinoma cells. *Proc Natl Acad Sci U S A*, 79 (9): 2845-2849.

Punnoose, E. A., Atwal, S., Liu, W., Raja, R., Fine, B. M., Hughes, B. G., Hicks, R. J., Hampton, G. M., Amler, L. C., Pirzkall, A. & Lackner, M. R. (2012) Evaluation of circulating tumor cells and circulating tumor DNA in non-small cell lung cancer: association with clinical endpoints in a phase II clinical trial of pertuzumab and erlotinib. *Clin Cancer Res*, 18 (8): 2391-2401.

Qiagen (2011) Therascreen EGFR Pyro Kit Handbook.

Qin, Z., Ljubimov, V. A., Zhou, C., Tong, Y. & Liang, J. (2016) Cell-free circulating tumor DNA in cancer. *Chin J Cancer*, 35 (1): 36.

Qiu, X., Zhang, J., Shi, W., Liu, S., Kang, M., Chu, H., Wu, D., Tong, N., Gong, W., Tao, G., Zhao, Q., Qiang, F., Zhu, H., Wu, Q., Wang, M. & Zhang, Z. (2016) Circulating MicroRNA-26a in Plasma and Its Potential Diagnostic Value in Gastric Cancer. *PLoS One*, 11 (3): e0151345.

Rajalingam, K., Schreck, R., Rapp, U. R. & Albert, S. (2007) Ras oncogenes and their downstream targets. *Biochim Biophys Acta*, 1773 (8): 1177-1195.

Computer ProgramRC, T. (2015) R: A language and environment for statistical computing.

Reiman, A., Kikuchi, H., Scocchia, D., Smith, P., Tsang, Y. W., Snead, D. & Cree, I. A. (2017) Validation of an NGS mutation detection panel for melanoma. *BMC Cancer*, 17 (1): 150.

Richards, M. A. (2009) The National Awareness and Early Diagnosis Initiative in England: assembling the evidence. *Br J Cancer*, 101 Suppl 2 S1-4.

Roche (2016) *Discover the COBAS EGFR Mutation Test v2*. [online] Available from: http://egfrmutationtestv2.roche.com/ (Accessed 23/08).

Rossi, D., Diop, F., Spaccarotella, E., Monti, S., Zanni, M., Rasi, S., Deambrogi, C., Spina, V., Bruscaggin, A., Favini, C., Serra, R., Ramponi, A., Boldorini, R., Foa, R. & Gaidano, G. (2017) Diffuse large B-cell lymphoma genotyping on the liquid biopsy. *Blood*,

Rosti, G., Bevilacqua, G., Bidoli, P., Portalone, L., Santo, A. & Genestreti, G. (2006) Small cell lung cancer. *Ann Oncol*, 17 Suppl 2 ii5-10.

Rubio, I., Wittig, U., Meyer, C., Heinze, R., Kadereit, D., Waldmann, H., Downward, J. & Wetzker, R. (1999) Farnesylation of Ras is important for the interaction with phosphoinositide 3-kinase gamma. *Eur J Biochem*, 266 (1): 70-82.

Russo, A., Ficili, B., Candido, S., Pezzino, F. M., Guarneri, C., Biondi, A., Travali, S., McCubrey, J. A., Spandidos, D. A. & Libra, M. (2014) Emerging targeted therapies for melanoma treatment (review). *Int J Oncol*, 45 (2): 516-524.

Sandulache, V. C., Williams, M. D., Lai, S. Y., Lu, C., William, W. N., Busaidy, N. L., Cote, G. J., Singh, R. R., Luthra, R. & Cabanillas, M. E. (2017) Real-Time Genomic Characterization Utilizing Circulating Cell-Free DNA in Patients with Anaplastic Thyroid Carcinoma. *Thyroid*, 27 (1): 81-87.

Schmid-Bindert, G., Wang, Y., Jiang, H., Sun, H., Henzler, T., Wang, H., Pilz, L. R., Ren, S. & Zhou, C. (2013) EBUS-TBNA Provides Highest RNA Yield for Multiple Biomarker Testing from Routinely Obtained Small Biopsies in Non-Small Cell Lung Cancer Patients - A Comparative Study of Three Different Minimal Invasive Sampling Methods. *PLoS One*, 8 (10): e77948.

Sepramaniam, S., Tan, J. R., Tan, K. S., DeSilva, D. A., Tavintharan, S., Woon, F. P., Wang, C. W., Yong, F. L., Karolina, D. S., Kaur, P., Liu, F. J., Lim, K. Y.,

Armugam, A. & Jeyaseelan, K. (2014) Circulating microRNAs as biomarkers of acute stroke. *Int J Mol Sci*, 15 (1): 1418-1432.

Shackelford, R. E., Whitling, N. A., McNab, P., Japa, S. & Coppola, D. (2012) KRAS Testing: A Tool for the Implementation of Personalized Medicine. *Genes Cancer*, 3 (7-8): 459-466.

Shapiro, B., Chakrabarty, M., Cohn, E. M. & Leon, S. A. (1983) Determination of circulating DNA levels in patients with benign or malignant gastrointestinal disease. *Cancer*, 51 (11): 2116-2120.

Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. (2007) Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer*, 7 (3): 169-181.

Shaw, J. A., Guttery, D. S., Hills, A., Fernandez-Garcia, D., Page, K., Rosales, B. M., Goddard, K. S., Hastings, R. K., Luo, J., Ogle, O., Woodley, L., Ali, S., Stebbing, J. & Coombes, R. C. (2017) Mutation Analysis of Cell-Free DNA and Single Circulating Tumor Cells in Metastatic Breast Cancer Patients with High Circulating Tumor Cell Counts. *Clin Cancer Res*, 23 (1): 88-96.

Shaw, J. A., Page, K., Blighe, K., Hava, N., Guttery, D., Ward, B., Brown, J., Ruangpratheep, C., Stebbing, J., Payne, R., Palmieri, C., Cleator, S., Walker, R. A. & Coombes, R. C. (2012) Genomic analysis of circulating cell-free DNA infers breast cancer dormancy. *Genome Res*, 22 (2): 220-231.

Shen, W. D., Chen, H. L. & Liu, P. F. (2014) EGFR gene copy number as a predictive biomarker for resistance to anti-EGFR monoclonal antibodies in metastatic colorectal cancer treatment: a meta-analysis. *Chin J Cancer Res*, 26 (1): 59-71.

Sherwood, J. L., Corcoran, C., Brown, H., Sharpe, A. D., Musilova, M. & Kohlmann, A. (2016) Optimised Pre-Analytical Methods Improve KRAS Mutation Detection in Circulating Tumour DNA (ctDNA) from Patients with Non-Small Cell Lung Cancer (NSCLC). *PLoS One*, 11 (2): e0150197.

Shields, J. M., Pruitt, K., McFall, A., Shaub, A. & Der, C. J. (2000) Understanding Ras: 'it ain't over 'til it's over'. *Trends Cell Biol*, 10 (4): 147-154.

Sholl, L. M., Xiao, Y., Joshi, V., Yeap, B. Y., Cioffredi, L. A., Jackman, D. M., Lee, C., Janne, P. A. & Lindeman, N. I. (2010) EGFR mutation is a better predictor of response to tyrosine kinase inhibitors in non-small cell lung carcinoma than FISH, CISH, and immunohistochemistry. *Am J Clin Pathol*, 133 (6): 922-934.

Sigdel, T. K., Vitalone, M. J., Tran, T. Q., Dai, H., Hsieh, S. C., Salvatierra, O. & Sarwal, M. M. (2013) A rapid noninvasive assay for the detection of renal transplant injury. *Transplantation*, 96 (1): 97-101.

Simbolo, M., Gottardi, M., Corbo, V., Fassan, M., Mafficini, A., Malpeli, G., Lawlor, R. T. & Scarpa, A. (2013) DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS One*, 8 (6): e62692.

Simpson, J. L. (2013) Cell-free fetal DNA and maternal serum analytes for monitoring embryonic and fetal status. *Fertil Steril*, 99 (4): 1124-1134.

Smouse, J. H., Cibas, E. S., Janne, P. A., Joshi, V. A., Zou, K. H. & Lindeman, N. I. (2009) EGFR mutations are detected comparably in cytologic and surgical pathology specimens of nonsmall cell lung cancer. *Cancer*, 117 (1): 67-72.

Sorber, L., Zwaenepoel, K., Deschoolmeester, V., Roeyen, G., Lardon, F., Rolfo, C. & Pauwels, P. (2017) A Comparison of Cell-Free DNA Isolation Kits: Isolation and Quantification of Cell-Free DNA in Plasma. *J Mol Diagn*, 19 (1): 162-168.

Sorich, M. J., Wiese, M. D., Rowland, A., Kichenadasse, G., McKinnon, R. A. & Karapetis, C. S. (2015) Extended RAS mutations and anti-EGFR monoclonal antibody survival benefit in metastatic colorectal cancer: a meta-analysis of randomized, controlled trials. *Ann Oncol*, 26 (1): 13-21.

Spindler, K. G. (2017) Methodological, biological and clinical aspects of circulating free DNA in metastatic colorectal cancer. *Acta Oncol*, 56 (1): 7-16.

Sriram, K. B., Tan, M. E., Savarimuthu, S. M., Wright, C. M., Relan, V., Stockwell, R. E., Clarke, B. E., Duhig, E. E., Yang, I. A., Bowman, R. V. & Fong, K. M. (2011) Screening for activating EGFR mutations in surgically resected nonsmall cell lung cancer. *Eur Respir J*, 38 (4): 903-910.

Su, K. Y., Kao, J. T., Ho, B. C., Chen, H. Y., Chang, G. C., Ho, C. C. & Yu, S. L. (2016) Implementation and Quality Control of Lung Cancer EGFR Genetic Testing by MALDI-TOF Mass Spectrometry in Taiwan Clinical Practice. *Sci Rep*, 6 30944.

Su, Z., Dias-Santagata, D., Duke, M., Hutchinson, K., Lin, Y. L., Borger, D. R., Chung, C. H., Massion, P. P., Vnencak-Jones, C. L., Iafrate, A. J. & Pao, W. (2011) A platform for rapid detection of multiple oncogenic mutations with relevance to targeted therapy in non-small-cell lung cancer. *J Mol Diagn*, 13 (1): 74-84.

Subramanian, J. & Govindan, R. (2007) Lung cancer in never smokers: a review. *J Clin Oncol*, 25 (5): 561-570.

Szpechcinski, A., Chorostowska-Wynimko, J., Kupis, W., Maszkowska-Kopij, K., Dancewicz, M., Kowalewski, J. & Orlowski, T. (2012) Quantitative analysis of free-circulating DNA in plasma of patients with resectable NSCLC. *Expert Opin Biol Ther*, 12 Suppl 1 S3-9.

Szpechcinski, A., Chorostowska-Wynimko, J., Struniawski, R., Kupis, W., Rudzinski, P., Langfort, R., Puscinska, E., Bielen, P., Sliwinski, P. & Orlowski, T. (2015) Cell-free DNA levels in plasma of patients with non-small-cell lung cancer and inflammatory lung disease. *Br J Cancer*, 113 (3): 476-483.

Tan, C. & Du, X. (2012) KRAS mutation testing in metastatic colorectal cancer. *World J Gastroenterol*, 18 (37): 5171-5180.

Taniguchi, K., Uchida, J., Nishino, K., Kumagai, T., Okuyama, T., Okami, J., Higashiyama, M., Kodama, K., Imamura, F. & Kato, K. (2011) Quantitative detection of EGFR mutations in circulating tumor DNA derived from lung adenocarcinomas. *Clin Cancer Res*, 17 (24): 7808-7815.

Thelwell, N., Millington, S., Solinas, A., Booth, J., Brown, T. (2000) Mode of action and application of Scorpion primers to mutation detection. *Nucleic Acids Res,* 1;28 (19): 3752-61.

ThermoFisher (2012) *SNaPshot® Multiplex System for SNP genotyping*. [online] Available from: https://tools.thermofisher.com/content/sfs/brochures/cms_101014.pdf (Accessed 18/01).

ThermoFisher (2016) *Single-Strand Conformation Polymorphism (SSCP) Analysis*. [online] Available from: https://www.thermofisher.com/uk/en/home/life-science/sequencing/fragment-analysis/single-strand-conformation-polymorphism-sscp.html# (Accessed 13/01/2017).

Thijssen, M. A., Swinkels, D. W., Ruers, T. J. & de Kok, J. B. (2002) Difference between free circulating plasma and serum DNA in patients with colorectal liver metastases. *Anticancer Res*, 22 (1A): 421-425.

Toh, C. K., Wong, E. H., Lim, W. T., Leong, S. S., Fong, K. W., Wee, J. & Tan, E. H. (2004) The impact of smoking status on the behavior and survival outcome of patients with advanced non-small cell lung cancer: a retrospective analysis. *Chest*, 126 (6): 1750-1756.

Tops, B., Normanno, N., Kurth, H., Amato, E., Mafficini, A., Rieber, N., Le Corre, D., Rachiglio, A., Reiman, A., Sheils, O., Noppen, C., Lacroix, L., Cree, I. A., Scarpa, A., Ligtenberg, M. & Laurent-Puig, P. (2015) Development of a semi-conductor sequencing-based panel for genotyping of colon and lung cancer by the Onconetwork consortium. *BMC Cancer*, 15 (1): 26.

Toro, P. V., Erlanger, B., Beaver, J. A., Cochran, R. L., VanDenBerg, D. A., Yakim, E., Cravero, K., Chu, D., Zabransky, D. J., Wong, H. Y., Croessmann, S., Parsons, H., Hurley, P. J., Lauring, J. & Park, B. H. (2015) Comparison of cell stabilizing blood collection tubes for circulating plasma tumor DNA. *Clin Biochem*, 48 (15): 993-998.

Vanni, I., Coco, S., Truini, A., Rusmini, M., Dal Bello, M. G., Alama, A., Banelli, B., Mora, M., Rijavec, E., Barletta, G., Genova, C., Biello, F., Maggioni, C. & Grossi, F. (2015) Next-Generation Sequencing Workflow for NSCLC Critical Samples Using a Targeted Sequencing Approach by Ion Torrent PGM Platform. *Int J Mol Sci*, 16 (12): 28765-28782.

Vliegen, L., Dooms, C., De Kelver, W., Verbeken, E., Vansteenkiste, J. & Vandenberghe, P. (2015) Validation of a locked nucleic acid based wild-type blocking PCR for the detection of EGFR exon 18/19 mutations. *Diagn Pathol*, 10 57.

Vojtek, A. B. & Der, C. J. (1998) Increasing complexity of the Ras signaling pathway. *J Biol Chem*, 273 (32): 19925-19928.

Volik, S., Alcaide, M., Morin, R. D. & Collins, C. C. (2016) Cell-free DNA (cfDNA): clinical significance and utility in cancer shaped by emerging technologies. *Mol Cancer Res*,

Wang, J., Han, X. & Sun, Y. (2017) DNA methylation signatures in circulating cell-free DNA as biomarkers for the early detection of cancer. *Sci China Life Sci*,

Warnecke-Eberz, U., Metzger, R., Holscher, A. H., Drebber, U. & Bollschweiler, E. (2016) Diagnostic marker signature for esophageal cancer from transcriptome analysis. *Tumour Biol*, 37 (5): 6349-6358.

Weber, B., Hager, H., Sorensen, B. S., McCulloch, T., Mellemgaard, A., Khalil, A. A., Nexo, E. & Meldgaard, P. (2014a) EGFR mutation frequency and effectiveness of erlotinib: a prospective observational study in Danish patients with non-small cell lung cancer. *Lung Cancer*, 83 (2): 224-230.

Weber, B., Meldgaard, P., Hager, H., Wu, L., Wei, W., Tsai, J., Khalil, A., Nexo, E. & Sorensen, B. S. (2014b) Detection of EGFR mutations in plasma and biopsies from non-small cell lung cancer patients by allele-specific PCR assays. *BMC Cancer*, 14 294.

Wei, H., Yuan, Y., Liu, S., Wang, C., Yang, F., Lu, Z., Wang, C., Deng, H., Zhao, J., Shen, Y., Zhang, C., Yu, X. & Xu, Q. (2015) Detection of circulating miRNA levels in schizophrenia. *Am J Psychiatry*, 172 (11): 1141-1147.

Williams, C., Ponten, F., Moberg, C., Soderkvist, P., Uhlen, M., Ponten, J., Sitbon, G. & Lundeberg, J. (1999) A high frequency of sequence alterations is due to formalin fixation of archival specimens. *Am J Pathol*, 155 (5): 1467-1471.

Willmore-Payne, C., Holden, J. A. & Layfield, L. J. (2006) Detection of EGFR- and HER2-activating mutations in squamous cell carcinoma involving the head and neck. *Mod Pathol*, 19 (5): 634-640.

Woo, S. M., Kim, M. K., Joo, J., Yoon, K. A., Park, B., Park, S. J., Han, S. S., Lee, J. H., Hong, E. K., Kim, Y. H., Moon, H., Kong, S. Y., Kim, T. H. & Lee, W. J. (2017) Induction Chemotherapy with Gemcitabine and Cisplatin Followed by Simultaneous Integrated Boost-Intensity Modulated Radiotherapy with Concurrent Gemcitabine for Locally Advanced Unresectable Pancreatic Cancer: Results from a Feasibility Study. *Cancer Res Treat*,

Woo, T. H., Patel, B. K., Cinco, M., Smythe, L. D., Symonds, M. L., Norris, M. A. & Dohnt, M. F. (1998) Real-time homogeneous assay of rapid cycle polymerase chain reaction product for identification of Leptonema illini. *Anal Biochem*, 259 (1): 112-117.

Wood, K. & Luke, J. J. (2016) Optimal Use of BRAF Targeting Therapy in the Immunotherapy Era. *Curr Oncol Rep*, 18 (11): 67.

Woolf, D. K., Slotman, B. J. & Faivre-Finn, C. (2016) The Current Role of Radiotherapy in the Treatment of Small Cell Lung Cancer. *Clin Oncol (R Coll Radiol)*,

Wu, Y. L., Sequist, L. V., Hu, C. P., Feng, J., Lu, S., Huang, Y., Li, W., Hou, M., Schuler, M., Mok, T., Yamamoto, N., O'Byrne, K., Hirsh, V., Gibson, N., Massey, D., Kim, M. & Yang, J. C. (2017) EGFR mutation detection in circulating cell-free DNA of lung adenocarcinoma patients: analysis of LUX-Lung 3 and 6. *Br J Cancer*, 116 (2): 175-185.

Xie, G., Xie, F., Wu, P., Yuan, X., Ma, Y., Xu, Y., Li, L., Xu, L., Yang, M. & Shen, L. (2015) The mutation rates of EGFR in non-small cell lung cancer and KRAS in colorectal cancer of Chinese patients as detected by pyrosequencing using a novel dispensation order. *J Exp Clin Cancer Res*, 34 63.

Xu, F., Wu, J., Xue, C., Zhao, Y., Jiang, W., Lin, L., Wu, X., Lu, Y., Bai, H., Xu, J., Zhu, G. & Zhang, L. (2012) Comparison of different methods for detecting epidermal growth factor receptor mutations in peripheral blood and tumor tissue of non-small cell lung cancer as a predictor of response to gefitinib. *Onco Targets Ther*, 5 439-447.

Xue, X., Teare, M. D., Holen, I., Zhu, Y. M. & Woll, P. J. (2009) Optimizing the yield and utility of circulating cell-free DNA from plasma and serum. *Clin Chim Acta*, 404 (2): 100-104.

Yan, Z., Lambert, N. C., Ostensen, M., Adams, K. M., Guthrie, K. A. & Nelson, J. L. (2006) Prospective study of fetal DNA in serum and disease activity during pregnancy in women with inflammatory arthritis. *Arthritis Rheum*, 54 (7): 2069-2073.

Yatabe, Y., Hida, T., Horio, Y., Kosaka, T., Takahashi, T. & Mitsudomi, T. (2006) A rapid, sensitive assay to detect EGFR mutation in small biopsy specimens from lung cancer. *J Mol Diagn*, 8 (3): 335-341.

Young, E. C., Owens, M. M., Adebiyi, I., Bedenham, T., Butler, R., Callaway, J., Cranston, T., Crosby, C., Cree, I. A., Dutton, L., Faulkes, C., Faulkner, C., Howard, E., Knight, J., Huang, Y., Lavender, L., Lazarou, L. P., Liu, H., Mair, D., Milano, A., Sandell, S., Skinner, A., Wallace, A., Williams, M., Spivey, V., Goodall, J., Frampton, J., Ellard, S. & On behalf of the Clinical Molecular Genetics Society Scientific, S. (2013) A Comparison of Methods for EGFR Mutation Testing in Non-Small Cell Lung Cancer. *Diagn Mol Pathol*,

Yu, S. C., Lee, S. W., Jiang, P., Leung, T. Y., Chan, K. C., Chiu, R. W. & Lo, Y. M. (2013) High-resolution profiling of fetal DNA clearance from maternal plasma by massively parallel sequencing. *Clin Chem*, 59 (8): 1228-1237.

Yuan, H., Zhu, Z. Z., Lu, Y., Liu, F., Zhang, W., Huang, G., Zhu, G. & Jiang, B. (2012) A modified extraction method of circulating free DNA for epidermal growth factor receptor mutation analysis. *Yonsei Med J*, 53 (1): 132-137.

Zhang, B. O., Xu, C. W., Shao, Y., Wang, H. T., Wu, Y. F., Song, Y. Y., Li, X. B., Zhang, Z., Wang, W. J., Li, L. Q. & Cai, C. L. (2015a) Comparison of droplet digital PCR and conventional quantitative PCR for measuring EGFR gene mutation. *Exp Ther Med*, 9 (4): 1383-1388.

Zhang, J. W., Qin, T., Hong, S. D., Zhang, J., Fang, W. F., Zhao, Y. Y., Yang, Y. P., Xue, C., Huang, Y., Zhao, H. Y., Ma, Y. X., Hu, Z. H., Huang, P. Y. & Zhang, L. (2015b) Multiple oncogenic mutations related to targeted therapy in nasopharyngeal carcinoma. *Chin J Cancer*, 34 (4): 177-183.

Zhang, S., Shen, H., Shu, X., Peng, Q. & Wang, G. (2017) Abnormally increased low-density granulocytes in peripheral blood mononuclear cells are associated with interstitial lung disease in dermatomyositis. *Mod Rheumatol*, 27 (1): 122-129.

Zhao, X., Han, R. B., Zhao, J., Wang, J., Yang, F., Zhong, W., Zhang, L., Li, L. Y. & Wang, M. Z. (2013) Comparison of epidermal growth factor receptor mutation statuses in tissue and plasma in stage I-IV non-small cell lung cancer patients. *Respiration*, 85 (2): 119-125.

Zheng, H., Zhang, L., Zhao, Y., Yang, D., Song, F., Wen, Y., Hao, Q., Hu, Z., Zhang, W. & Chen, K. (2013) Plasma miRNAs as diagnostic and prognostic biomarkers for ovarian cancer. *PLoS One*, 8 (11): e77853.

Zhu, C., Ren, C., Han, J., Ding, Y., Du, J., Dai, N., Dai, J., Ma, H., Hu, Z., Shen, H., Xu, Y. & Jin, G. (2014) A five-microRNA panel in plasma was identified as potential biomarker for early detection of gastric cancer. *Br J Cancer*, 110 (9): 2291-2299.

Zhuang, L., Xu, L., Wang, P. & Meng, Z. (2015) Serum miR-128-2 serves as a prognostic marker for patients with hepatocellular carcinoma. *PLoS One*, 10 (2): e0117274.

Zsikla, V., Baumann, M. & Cathomas, G. (2004) Effect of buffered formalin on amplification of DNA from paraffin wax embedded small biopsies using real-time PCR. *J Clin Pathol*, 57 (6): 654-656.